2018

Structural underpinnings of membrane association and mechanism in the monotopic phosphoglycosyl transferase superfamily

BOSTON UNIVERSITY

SCHOOL OF MEDICINE

Dissertation

STRUCTURAL UNDERPINNINGS OF MEMBRANE ASSOCIATION AND MECHANISM IN THE MONOTOPIC PHOSPHOGLYCOSYL TRANSFERASE SUPERFAMILY

by

LEAH CHRISTINE RAY

B.S., The University of Texas at Austin, 2012
B.S., The University of Texas at Austin, 2012

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2018

Approved by

First Reader _____

Karen N. Allen, Ph.D.
Professor of Chemistry

Second Reader _____

Barbara Imperiali, Ph.D.
Class of 1922 Professor of Biology and Chemistry
Massachusetts Institute of Technology

"Wilt thou not chase the white whale?" – Captain Ahab

DEDICATION

I would like to dedicate this work to my husband, Matt, and my sister, Rachel, who have been my biggest supporters and role models in this endeavor.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor Dr. Karen Allen for her invaluable support and mentorship through my time in graduate school. Her passion and zeal for the scientific questions we were trying to answer, were often just the encouragement I needed to keep tackling some very tough questions. Without her ceaseless encouragement, I'm sure I would have given up after the first 100, let alone 700+, crystals of PglC I screened to no avail. Her optimism and encouragement helped me climb the ladders and celebrate the chutes of my Ph.D. Her mentorship has made me a confident scientist, ready and willing to address difficult questions in the future. And like the T-Rex with the grabber arm says: "I am... unstoppable", I appreciate her making me actually believe that.

I am very appreciative to have had the opportunity to work in collaboration with Dr. Barbara Imperiali to solve the structure of the troublesome little protein PglC. Her expertise and intense passion for science encouraged me to think critically about the even the smallest of problems. I am so grateful to have been a part of Team PglC.

I must acknowledge Team PglC and Team Nanodiscs for being a wonderful collaborative group. I am very proud of the scientific discoveries we have all made together. Even the simplest "Go PglC!" at the end emails reminded me in tough

and discouraging moments, that we were all working towards the same goal. I was never met with anything other than the greatest encouragement and best trouble-shooting advise from everyone. I'm thankful for the opportunity to work with such thoughtful and determined scientists. I am especially grateful to Debasis for his expert preparation of all of the protein I used in growing crystal after crystal of PglC. I am so happy our hard work finally paid off. Working closely with Sonya on the PglC structure, all things nanodiscs along with JM, and our review of monotopic proteins has been an utmost pleasure. And, at this point it is probably necessary for me to acknowledge Clover in Kendall Square for the best rosemary fries and sandwiches to fuel my many trips back and forth across the Mass Ave Bridge toting a precious styrofoam box.

I would like to thank Dr. Lee Makowski, not only for serving on my committee, but also for his openness and willingness to lend myself and Karen his expertise in all things SAXS. Being able to discuss my SAXS research with an expert gave me great confidence in learning a challenging new method.

I wish to thank my committee chair, Dr. Ben Wolozin, and Dr. Shelley Russek for their continued support and encouragement throughout my time in the Pharmacology department. Their assistance in serving on both my qualifying examination and dissertation advisory committees is greatly appreciated.

The Allen Lab has been a wonderful place to work, leaving me with so many colleagues to acknowledge. I am most thankful to all Allenites, old and new, for creating such a supportive, collaborative, and wonderfully unique scientific environment. Without everyone, I'm sure I would have quit graduate school long before this point. I am thankful to the Bridge Club for making the sometimes arduous days full of laughter, encouragement, thoughtful discussion, and of course, desk beers. I am especially grateful to Kath3rin3 for h3r fri3ndship; for our mutual appr3ciation for sci3nc3 idols, biophysics pap3rs, Clov3r lunch3s, and punny pap3r titl3s; and mostly for putting up with my inc3escant tortrur3 of h3r with odd numb3rs, 3sp3cially thr33s (333). I am grateful to Margarita for always inspiring everyone in the lab with your passion for doing great science, especially when it comes to all things flavin. I wish her luck in the building of the army of intense mini Margaritas, and look forward to all of the great things she will do. I am thankful to JP for encouraging me through (*too*) optimistic graduation timelines (I'm only about 3 months off!), for teaching me the importance of a thoughtful well-put-together presentation, for always playing devil's advocate to keep me grounded, and for ultimately making me a better critical thinker and scientist. I am grateful to Christine (Teeny Weeny Christineey/ the Shonda to my Vonda) for our lipin/ dullard commiseration, for keeping me company on many a

late night beam-time (especially with all 8 HPs in a row), and for being a constant, colorful reminder to have fun despite all of the stress. Always remember: “It’s a new day! Everything is wonderful!”.

I also wish acknowledge the BU Chemistry department for being my adoptive home and for always treating me as one of their own even though I am an impostor. And to acknowledge my actual home, I wish to thank Dan, Neema, and Trevor for being such great friends and colleagues when we first joined the Pharmacology department together in 2012. I am very grateful for the opportunity to have spent many hours in our early years of graduate school with them.

Lastly, I am eternally grateful for the love and support of my entire family near and far. My husband, Matt, was always there to celebrate my successes and to lend a supportive ear to my exhausted complaints about rough days (years) in the lab. I am looking forward to our next adventure in San Francisco. I very am grateful for my parents, and especially for my sister, Rachel, for always being so supportive of any endeavor I undertake.

STRUCTURAL UNDERPINNINGS OF MEMBRANE ASSOCIATION AND MECHANISM IN THE MONOTOPIC PHOSPHOGLYCOSYL TRANSFERASE SUPERFAMILY

LEAH CHRISTINE RAY

Boston University School of Medicine, 2018

Major Professor: Karen N. Allen, Ph.D., Professor of Chemistry

ABSTRACT

In prokaryotes, protein glycosylation can be a determinant of pathogenicity as it plays a role in host adherence, invasion, and colonization. Impairment of glycosylation in some organisms, for example N-linked glycosylation in *Campylobacter jejuni,* leads to decreased pathogenicity; thus, opening new avenues for the development of antivirulence agents. A member of the protein glycosylation (*pgl)* gene locus in *C. jejuni,* PglC, is predicted single-pass transmembrane (TM) protein, that catalyzes the phosphoglycosyl transferase (PGT) reaction in the first membrane-committed step of the N-linked glycosylation pathway. The small size of PglC (201 aa) compared to homologous PGTs suggests it may represent the minimal catalytic unit for the monotopic PGT superfamily.

Herein, the structure of *C. concisus* PglC including its putative TM domain has been solved to 2.74 Å resolution to reveal a novel protein fold with a unique alpha-helix-associated beta-hairpin (AHABh) motif and largely solvent-exposed structure. There is noted a parsimony of fold in the form of short-range motifs underpinning the structural basis for critical functions of PglC: membrane association and active-site geometry. Biochemical and bioinformatics studies support structural evidence suggesting the crystallographically-observed, kinked TM helix is re-entrant on the cytoplasmic face of the membrane rather than membrane spanning. Thus, PglC represents a first-in-class structure of a novel membrane interaction mode for monotopic membrane proteins. Additionally, the AHABh-motif and active-site helical geometry establishes co-facial positioning of the catalytic-dyad. Molecular docking of PglC substrates, undecaprenyl phosphate (UndP) and UDP-N,N-diacetylbacillosamine (UDP-diNAcBac), within the active-site reveals co-incident binding sites, consistent with the proposed ping-pong enzymatic mechanism.

Loading of PglC into membrane-bilayer nanodiscs (ND) allows for the investigation of PglC structure and function within a native-like membrane environment by small-angle x-ray scattering (SAXS). Observation of PglC in ND via SAXS confirms the application of the method for studying small, integral, monotopic

membrane proteins in a membrane environment. Moreover, development of a mathematical approach by which resident-protein: ND stoichiometry can be deduced from measured scattering intensity enables independent confirmation of loading stoichiometry.

Overall, the membrane-interaction modality observed for PglC is the first structurally characterized example of a new membrane association mode for monotopic proteins with the membrane. These studies provide insight into the structural determinants of the chemical mechanism and substrate-binding for *C. concisus* PglC and for the extensive homologous monotopic PGT superfamily, thus allow homology modeling and enabling future inhibitor design.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER TWO

CHAPTER THREE

CHAPTER FOUR

CHAPTER FIVE

APPENDIX A1

APPENDIX A2

LIST OF FIGURES

CHAPTER ONE

CHAPTER TWO

CHAPTER THREE

CHAPTER FOUR

CHAPTER FIVE

CHAPTER SIX

APPENDIX A1

APPENDIX A2

LIST OF ABBREVIATIONS

Standard 3-letter and 1-letter codes are used for the 20 natural amino acids.

aa ..... amino acid

AH ..... amphipathic helix

AHABh ..... α-helix-associated β-hairpin

APBS ..... Adaptive Poisson-Boltzmann Solver

ASAXS ..... anomalous small angle x-ray scattering

ASU ..... asymmetric unit

AUC ..... analytical ultracentrifugation

BNL ..... Brookhaven National Lab

BU ..... Boston University

BUMC ..... Boston University Medical Center

C. concisus ..... *Campylobacter concisus*

C. jejuni ..... *Campylobacter jejuni*

CC ..... correlation coefficient

CoM ..... center of mass

CPS ..... capsular polysaccharide

Da ..... Daltons

DAG ..... diacylglycerol

DDM.....dodecylmaltoside

diNAcBac ..... N'N-diacetylbacillosamine

DLS.....dynamic light scattering

DMPC 1,2-dimyristoyl-sn-glycero-3-phosphocholine

DMPG 1,2-dimyristoyl-sn-glycero-3-phosphoglycerol

DOPC 1,2-Dioleoyl-sn-glycero-3-phosphocholine

DOPS.....1,2-Dioleoyl-sn-glycero-3-phosphoserine

DTT.....dithiothreitol

E. coli .....*Escherichia coli*

FAAH..... fatty acid acyl hydrolase

FPLC.....fast protein liquid chromatography

Gal-P.....Galactose-phosphate

GalNac .....N-acetylglucosamine

GT ..... glycosyltransferase

H. pullorum ..... *Helicobacter pullorum*

HAD ..... haloalkanoic acid dehalogenase

HTS.....high-throughput screening

HWI.....Haputmann-Woodward Institute

IPTG .....isopropyl-1-β -D-1-galactopyranoside

L. geestiana ..... *Leigonella geestiana*

LB ..... Luria Broth

LBT ..... lanthanide binding tag

LC-MS ..... liquid chromatography mass spectrometry

LPS ..... lipopolysaccharide

M. smegmatis ..... *Mycobacterium smegmatis*

MALDI-MS..... matrix assisted laser desorption and ionization mass spectrometry

MD ..... molecular dynamics

MIT ..... Massachusetts Institute of Technology

MM ..... molecular mass

MR ..... molecular replacement

MSP ..... membrane scaffold protein

MST ..... microscale thermophoresis

N. gonorrhoeae ..... *Neisseria gonorrhoeae*

N. meningitidis ..... *Neisseria meningitidis*

NCS ..... Noncrystallographic Symmetry

ND ..... nanodiscs

NLIP ..... N-terminal domain of lipin

NSLS ..... National Synchrotron Light Source

OD ..... optical density

P. pudita..... *Pseudomonas pudita*

PA .....phosphatidic acid

PDB..... Protein Data Bank

PE..... phosphotidyl ethanol

PG ..... phosphotidyl glycerol

PEG.....Polyethylene glycol

PGHS.....prostaglandin H2 synthase

Pgl.....protein glycosylation

PGT.....phosphoglycosyltransferase

PO_4^{3-} ..... inorganic phosphate

POPE 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine

POPG 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoglycerol

PPM Server..... Positioning of Proteins in Membrane

Pren-P..... polyprenol phosphate

R_c.....cross-sectional radius of gyration

R_g.....radius of gyration

RMH.....reentrant membrane helix

RMSD ..... Root Mean Square Deviation

S-SAD ..... sulfur- single anomalous dispersion

S-synthase ..... squalene synthase

S. aureus ..... *Staphylococcus aureus*

S. enterica ..... *Salmonella enterica*

S. pneumoniae ..... *Streptococcus pneumoniae*

SAD ..... single anomalous dispersion

SAS ..... solvent-accessible surface area

SAXS ..... small angle x-ray scattering

SCAM ..... substituted cysteine accessibility method

SE ..... sedimentation equilibrium

SEC-SAXS ..... size-exclusion chromatography coupled SAXS

SeMet ..... Selenomethionine

SEV ..... solvent-excluded volume

SF ..... superfamily

SGSG ..... Serine-Glycine-Serine-Glycine

SV ..... sedimentation velocity

SVR ..... surface-to-volume ratio

TLC ..... thin layer chromatography

TLS ..... Translation-Libration-Screw

TM .....transmembrane

UDP ..... uridine diphosphate

UDP-diNAcBac..... uridine diphosphate- N'N'-diacetylbacillosamine

UDP-GlcNAc .....uridine diphosphate- N-acetylgulcosamine

UMP .....uridine monophosphate

UndP ..... undecaprenol phosphate

UV..... ultraviolet

WT .....wild type

X. campestris ..... *Xanthomonas campestris*

XAS.....x-ray absorption spectra

CHAPTER ONE

Introduction

1.1 Glycoconjugates impart structure and function in bacteria

Glycosylation of proteins is a commonly occurring post-translational modification. In general, glycosylation pathways involve monosaccharides that are linked by glycosidic bonds to create glycan chains (1). These glycan chains are attached to serine (Ser), threonine (Thr), or asparagine (Asn) residues of a target protein. The residue that is modified determines the identity of the glycosylation as either nitrogen-linked (N-linked), if on Asn, or oxygen-linked (O-linked), if on Ser or Thr. Notably, 30 - 40% of the eukaryotic proteome is glycosylated to produce glycoproteins or other downstream glycoconjuates, which play roles in critical cellular processes including protein folding, protein trafficking, cell-cell interactions, and the host-immune response (2, 3). Whereas the prevalence of glycosylation is significantly higher in eukaryotes, glycosylation plays a no-less critical role for members of the prokaryotic kingdom. The resulting glycoproteins and glycoconjugates in bacteria are involved in pathways promoting improved survival and pathogenicity where they play structural and modulatory roles through the formation of cell wall peptidoglycan and mediation of host-cell adhesion, invasion, and colonization.

The prevalence of glycoconjugates within critical cellular processes necessitates the extensive characterization of bacterial glycosylation pathways. In bacteria, glycosylation occur via two main biosynthetic mechanisms, sequential transfer or *en bloc* transfer (4). Sequential glycosylation mechanisms involve transfer of sugar moieties directly to the substrate (lipid, protein, *etc*.) of interest by glycosyltransferases (GT) enzymes to build up simple glycans. In contrast, *en bloc* bacterial glycosylation pathways involve commencement of glycosylation through a phopshoglycosyl transferase (PGT) followed series of GT enzymes which function at the cytoplasmic membrane interface to create and grow an oligosaccharide chain linked to a membrane-embedded polyprenol-phosphate (Pren-P). Upon translocation of the glycan chain to the periplasmic space the fully-assembled glycan chains can then be transferred *en bloc* to a protein target or further processed into glycoconjugates. *En bloc* glycosylation enables biosynthesis of complex glycoconjugates, such as peptidoglycan, capsular polysaccharides (CPS), O-antigen and lipopolysaccharides (LPS), and glycoproteins (5-7) (figure 1.1).

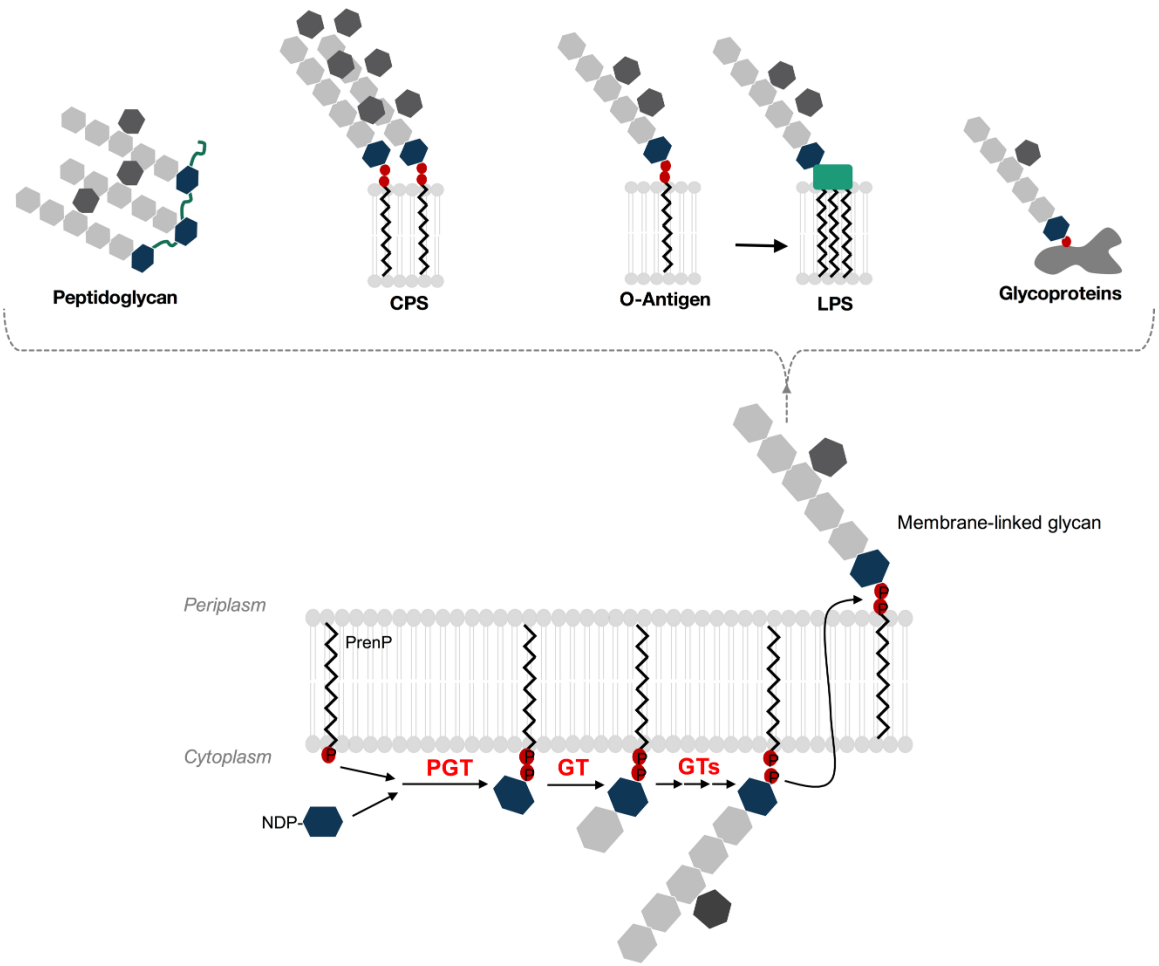

Figure 1.1. Exemplar downstream glycoconjugates arising from en bloc glycoconjugate biosynthetic pathways in bacteria. Phosphoglycosyl transferase (PGT) and glycosyltransferase (GT) enzymes comprising the en bloc biosynthetic pathways depicted in red text. CPS is capsular polysaccharide. LPS is lipopolysaccharide. PrenP is polyprenol-phosphate. NDP is nucleotide diphosphate. Green block represents lipid A anchor. Hexagons represent sugar moieties.

These downstream glycoconjugates produced by *en bloc* glycosylation pathways are associated with bacterial survival and pathogenicity. For example, peptidoglycan comprises the cell wall essential for bacterial viability and

protection. Cell-wall structural integrity, as well as host-pathogen interactions, are further mediated by the cell-wall polysachharide CPS. Together with CPS, O-antigen and LPS comprise the majority of the antigens on the surface of the Gram-negative bacterium *Eschericia coli* (8). Similarly, N- and O-linked glycoproteins resultant from *en bloc* glycosylation pathways have been implicated in host-cell adhesion, colonization, and invasion (9, 10). The roles played by bacterial glycoconjugates as determinants of bacterial pathogenicity necessitate a detailed understanding of both the functions glycoconjugates and the glycosylation pathways.

1.2 N-linked glycosylation in pathogenic bacteria

The N-linked glycosylation pathway from *Campylobacter jejuni* is the most widely characterized N-linked glycosylation pathway and often serves as the prototype for the study of bacterial *en bloc* glycosylation pathways *(6, 9).* Genetic impairment of the N-linked glycosylation pathway in *Campylobacter jejuni* via glycosylation-null mutants results in decreased pathogenicity via a decrease in the ability of a bacterium to adhere to, invade, and/or colonize its host (6, 9). From these studies, it is only clear that bacterial virulence requires N-linked glycosylation pathways as these mutations abrogate the formation of all

glycoproteins within the bacteria; therefore, making it uncertain if this effect is the result of a singular or cumulative loss of glycoproteins within the cells.

The N-linked glycosylation pathway in *C. jejuni* shares similarity to pathways in *Nesseria gonhorreae* and *Heliobacter pullorum,* all three of which are major causative agents of human disease. Enzymes which make up the *C. jejuni* N-linked glycosylation pathway, summarized in figure 1.2, are encoded by a single protein glycosylation (*pgl*) gene locus (11). The pathway starts with soluble UDP-GlcNAc (uridine diphosphate N-acetylglucosamine) which is then transformed into the unique sugar containing UDP-N,N′-diacetyl-bacillosamine (UDP-diNAcBac) by PglD, PglE, and PglF. PglC, an integral membrane PGT, then links this unique sugar-phosphate to a membrane-resident undecaprenol phosphate (UndP) and releases UMP. At this point the nascent glycan chain is committed to the membrane with the membrane-bound PrenP-PP-diNAcBac allowing for the commencement of glycoconjugate biosynthesis. Through the action of subsequent glycosyltransferases PglA, PglJ, PglH, and PglI, the heptasaccharide glycan chain is fully assembled. The PglK flippase translocates the completed glycan (GalNAc-α-1,4-GalNAc-α-1,4-(Glc-β-1,3)-GalNAc-α-1,4-GalNAc-α-1,4-GalNAc-α-1,3-diNAcBac) from the cytoplasmic space into the periplasmic space, where PglB can then transfer the glycan from the UndP moiety to Asn residues on acceptor

proteins (12, 13). Notably, the crystal structures of the *Campylobacter* oligiosaccharyl transferase PglB (PDB ID: 3RCE, *C. lari; 3AAG, C. jejuni*) and the flippase PglK (PDB ID: 5C76, *C. jejuni*) have been previously determined.

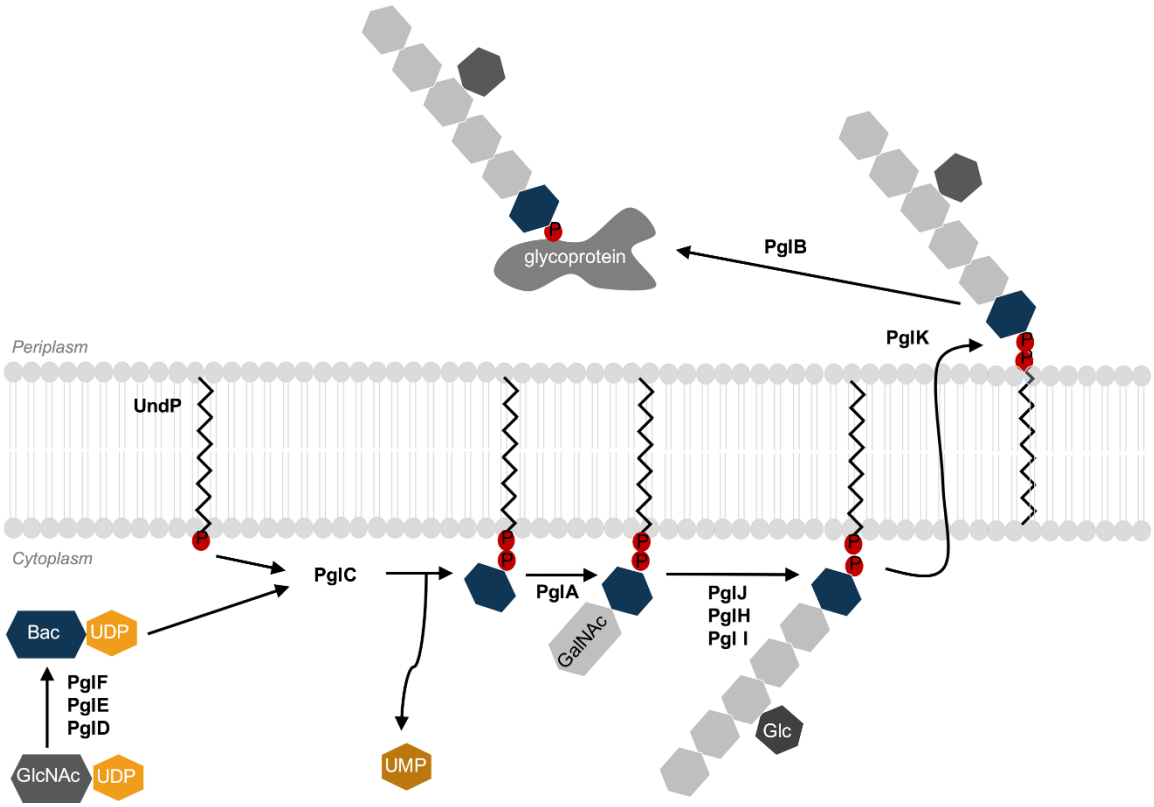

Figure 1.2. N-linked glycosylation pathway in C. jejuni.

Within *en bloc* glycosylation pathways, typified by the *C. jejuni* N-linked pathway, catalysis of the the first membrane-committed step by a membrane-bound phosphoglycosyl transferase (PGT) is a shared feature. The PGT isalso referred to as a priming or initiating GT, catalyzes the transfer of a of a C1'-

phosphosugar from a soluble nucleoside-diphosphate sugar to a polyprenol phosphate (PrenP) resident in the cytoplasmic face of the membrane (Figure 1.3) (14, 15). PGTs effectively act as gatekeepers of glycan biosynthesis within these pathways; as such, they are well-poised for inhibition. It should be noted that PGTs are not unique to bacteria; they are also found in similar eukaryotic glycosylation pathways, where they initate N-linked glycosylation through the transfer of a sugar to a dolichol linear polyprenol (16).

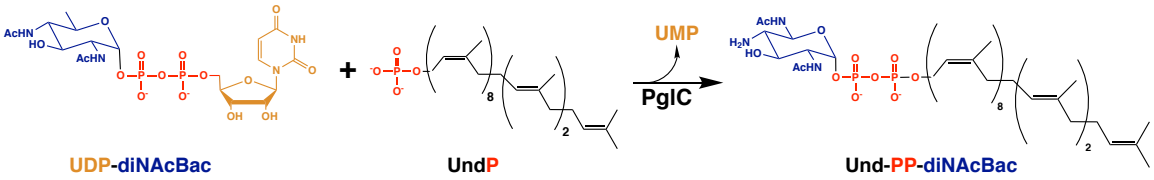

Figure 1.3. General PGT reaction scheme in *C. jejuni* PglC as an exemplar for bacterial PGT reactions.

1.3 Bacterial PGT enzyme superfamiles

The PGT reaction is carried out by two isofunctional superfamilies of PGT enzymes: the polytopic PGT superfamily and the monotopic PGT superfamily (14, 17). While the superfamilies catalyze the same general reaction, they are structurally and mechanistically divergent. The polytopic PGT superfamily, typified by the MraY and WecA subfamilies, is characterized by 10-11 transmembrane (TM) helices (Figure 1.4a) (14). MraY, composed of 10 TM helices

is responsible for the transfer of 1-phospho-MurNAc-pentapeptide (L-Ala-γ-D-Glu-diaminopimelic acid/ L-Lys-D-Ala-D-Ala) to UndP in the membrane to initiate the biosynthesis of peptidoglycan. Similarly, the 11-TM helix enzyme WecA catalyzes the transfer of GlcNAc-1-phosphate to PrenP (18) within the O-antigen biosynthesis pathway. Of the polytopic PGTs, MraY remains the only structurally characterized member of the superfamily (19). In contrast, the isofunctional monotopic PGT superfamily is predicted to assume single-pass transmembrane topology with a hydrophobic membrane-associated domain and a soluble globular domain. The monotopic PGT superfamily is subdivided into the PglC-monofunctional, PglB-bifunctional, and WbaP subfamilies (Figure 1.4b) (20). The monofunctional-PglC and bifunctional-PglB subfamilies catalyze the transfer of the unique sugar (diNAcBac) with the C1'-phosphate to UndP (20, 21) in the biosynthesis of N-linked and O-linked glycoproteins, respectively. The WbaP subfamily catalyzes the transfer of galactose-phosphate (Gal-P) to UndP (22) to initiate the O-antigen biosynthetic pathway. Prior to the work elucidating the molecular structure of *C. concisus* PglC presented in this dissertation, this entire superfamily of proteins had not been characterized experimentally in terms of three-dimensional structure.

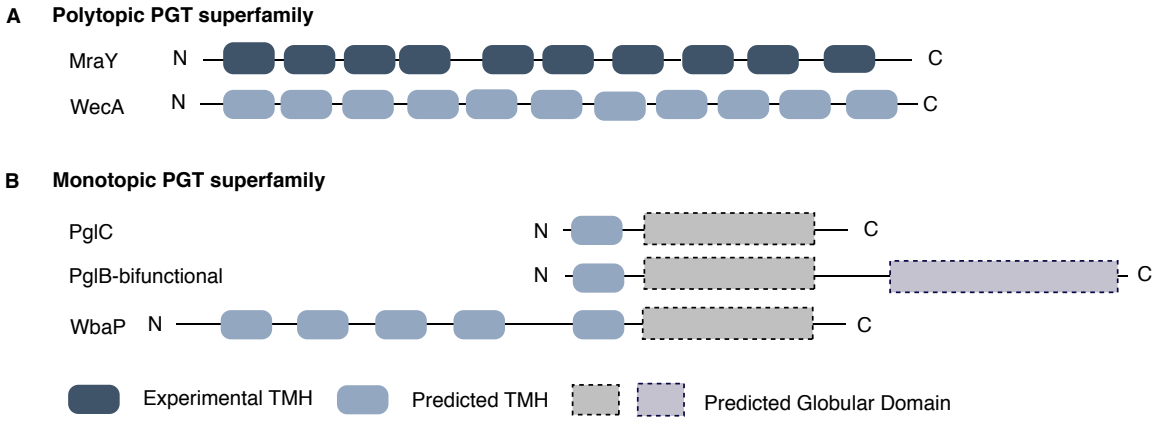

Figure 1.4. Schematic of topology differences between members of the polytopic and monotopic PGT superfamilies.

In addition to the stark differences in molecular architecture, the polytopic and monotopic PGT superfamiles are mechanistically divergent. Mechanistic studies of MraY and WecA have demonstrated that the PGT reaction proceeds through a ternary complex reaction mechanism; both substrates, the UDP-sugar and the polyprenol phosphate, occupy the active site simultaneously for catalysis (Figure 1.4) (23, 24). In contrast, recent biochemical studies of *C. concisus* PglC provide evidence that the monotopic PGTs proceeds through a covalent-phospho-sugar intermediate which undergoes two-step, ping-pong kinetics (Figure 1.5). In the first step, the UDP-sugar enters the active site where the enzyme nucleophile, the carboxylate of Asp, attacks at the β-phosphate of the UDP to create the phospho-sugar-enzyme intermediate and to liberate the by-product, UMP. In the

second step, the membrane-resident UndP enters the active site where it attacks the phospho-sugar intermediate to release the enzyme aspartate and produce the membrane-bound Und-PP-diNAcBac product (25). As structure and function are intimately married in biomolecules, it is unsurprising that the marked difference in molecular architecture is mirrored by a stark difference in enzymatic mechanism between the polytopic and monotopic PGT superfamilies.

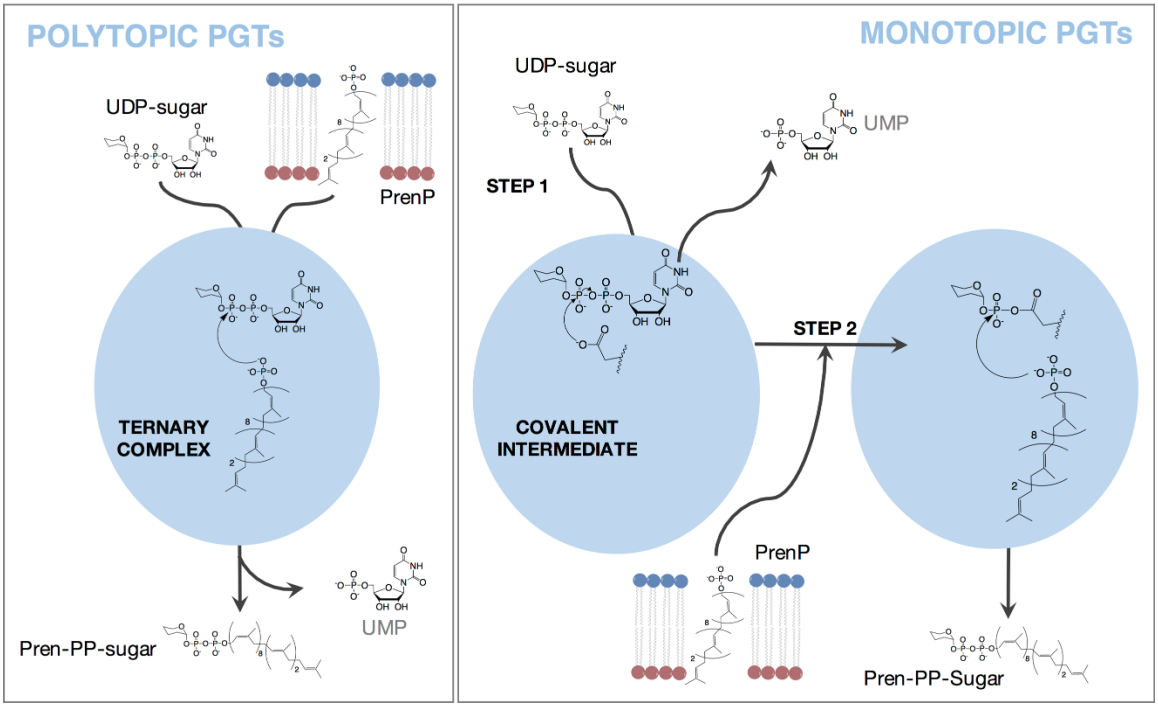

Figure 1.5. Schematic representation of polytopic and monotopic PGT enzymatic mechanisms. Enzyme active site represented by blue circle. PGT reaction byproduct, UMP, represented by gray text.

1.4 PGTs as targets for anti-virulence and antibacterial agent development

PGTs, in general, carry out the first membrane-committed step of many glycoconjugate pathways. It has been noted that PGT enzymes are inhibited by nucleoside antibiotics (26). Many of such antibiotics appear to be a bi-substrate analog inhibitors based on their chemical structures as they possess chemical moieties similar to the substrates (UDP-Sugar and PrenP) of the PGT enzymes. Tunicamycin, a member of this class of antibacterial compounds, possesses a nucleoside analog, sugar analog, Pren-P analog, and a pseudo-GalNAc (N-acetylglucosamine) moiety to serve as a diphosphate mimic (Figure 1.5) (27).

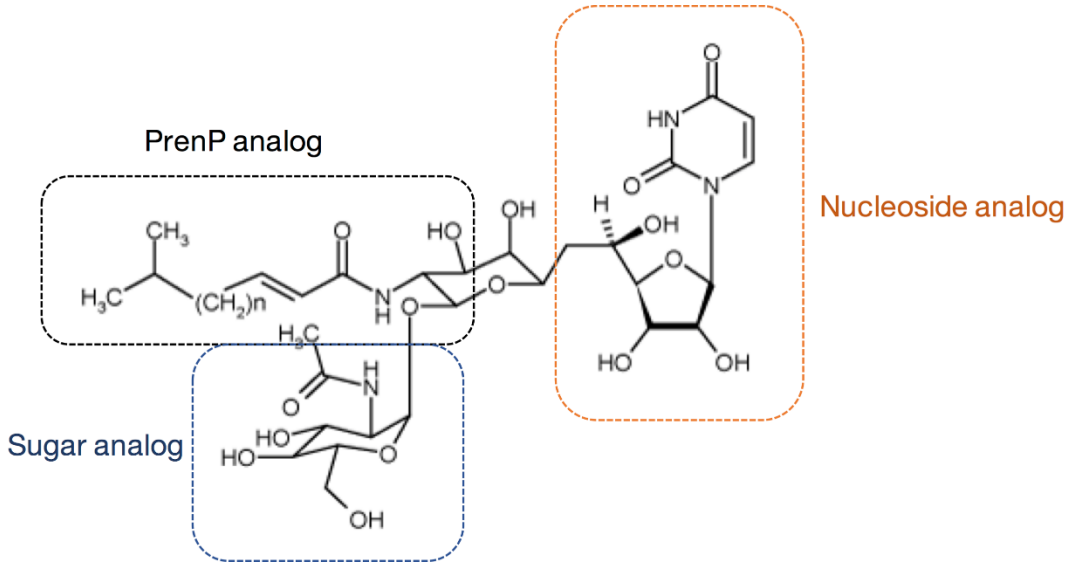

Figure 1.6. Chemical structure of the nucleoside antibiotic Tunicamycin. Analogous chemical groups highlighted.

As a result, tunicamycin is a potent inhibitor of both eukaryotic and prokaryotic glycosylation. Activity assays show that WecA from the polytopic PGT superfamily, has reduced activity upon inhibition by tunicamycin. Conversely, in the same assay *C. jejuni* PglC, from the monotopic PGT superfamily, activity was not inhibited (20). Given what is currently known about the differing mechanisms for the polytopic PGT family in comparison to the monotopic PGT family, it follows that the architecture of the enzyme active sites must be largely different. From this view it is unsurprising that the polytopic PGT WecA was inhibited by tunicamycin as the antibiotic chemical structure could functionally mimic the bi-substrate complex in the active site. Conversely, for the monotopic PglC substrate binding cannot be coincident as the covalent phospho-sugar intermediate mechanism suggests; therefore, the tunicamycin scaffold is no-longer a functional mimic for the catalytically-active enzyme-substrate complex. Overall, this differential sensitivity to tunicamycin suggests a detailed molecular understanding of the fundamental structural and mechanistic divergence of the monotopic and polytopic PGTs is necessary to provide an avenue for development of monotopic-PGT-specific inhibitors.

1.5 *C. jejuni* PglC represents the minimal catalytic unit for monotopic PGTs

Development of specific and potent inhibitors of N-linked glycosylation in bacteria necessitates a molecular understanding of the monotopic PGT superfamily. Thus, it is critical to have an experimentally determined three-dimensional structure of one of the superfamily members. As discussed earlier, the monotopic PGT superfamily is characterized by three subfamilies: the PglC-monofunctional, PglB-bifunctional, and WbaP subfamilies (17, 20). Differences in sugar specificity and molecular architecture between each subfamily account for their differentiation (Figure 1.6) (15). The PglC-monofunctional subfamily is structurally characterized by a single predicted TM helix domain and a small globular domain. The PglB-bifunctional subfamily contains a C-terminal-domain homologous to the PglC-monofunctional subfamily fused to a soluble N-terminal acetyltransferase domain homologous to the trimeric β-helix of PglD in *C. jejuni* (13). Seemingly misfit for the monotopic classification, the WbaP subfamily is characterized by a multipass transmembrane topology; however, truncation mutagenesis in a WbaP homolog, WcaJ, suggests that the C-terminal domain, comprised of one TM helix homologous to the PglC-monofunctional subfamily, is sufficient for full PGT activity making the family functionally monotopic (Figure 1.4, gray dashed boxes) (22). Over the entire superfamily, the PglC-

monofunctional subfamily, with its relatively small size (~200 aa) and small cytosolic and membrane-associated domains, represents the minimal catalytic unit for the monotopic PGT reaction.

In the absence of an experimentally-determined structure, Lukose *et al.* leveraged biochemical structure-function analyses and bioinformatics to construct structural model of the cytosolic globular domain of *C. jejuni* PglC (17). Ultimately, the model was calculated by employing three-dimensional distance constraints suggested by residue evolutionary-covariance across a monotopic PGT superfamily alignment of 15,000 sequences. The PglB-bifunctional and WbaP subfamilies share significant sequence identity with the PglC-monofunctional subfamily (53% and 34%, respectively); thus, any structural insight gained for the minimal PglC scaffold can be readily translated to the entire monotopic PGT family (17).

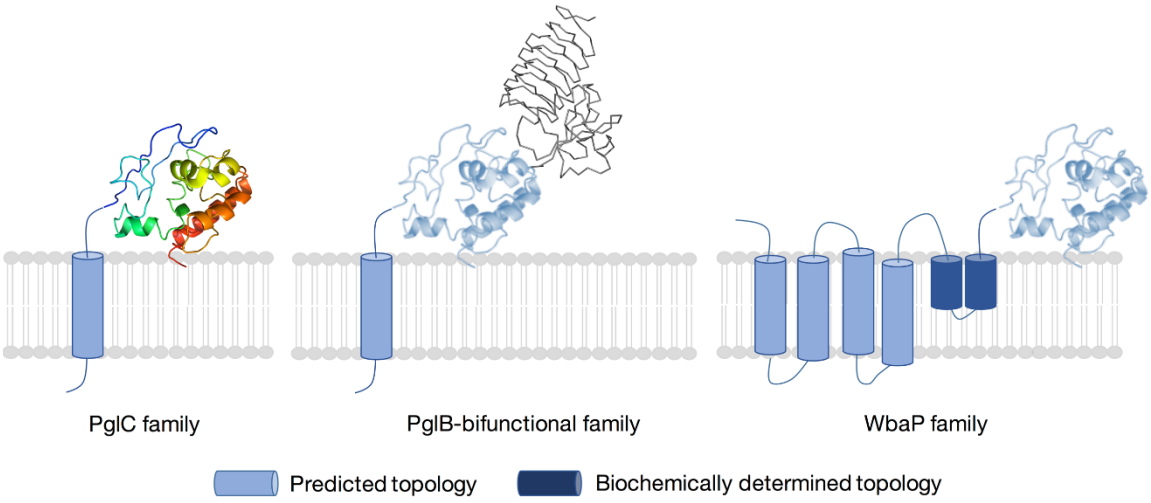

Figure 1.7. PglC represents the minimal catalytic unit for the monotopic PGT reaction. EV-fold model of globular domain of C. jejuni PglC (17) shown in color ramp from blue N-terminus to red C-terminus. Homologous soluble domain of the PglB-bifunctional and WbaP families represented by light blue PglC domain model (17). Experimentally determined acetyltransferase domain of PglB (4M98) from *Neisseria gonorrhoeae* shown in gray ribbon (28).

As a target for structure elucidation, the smallest (23 kDa, 200 aa) of the PglC-monofunctional subfamily members is from the well-characterized N-linked glycosylation pathway in *C. jejuni.* Overall, the enzyme PglC from *C. jejuni* presents as an optimal candidate for structure elucidation and inhibitor development as it represents the minimal catalytic unit for the superfamily and catalyzes the first membrane-committed step of the N-linked glycosylation pathway in the bacterium.

Not only is the structure of PglC of significant biological interest, the structure of PglC is poised to expand general knowledge in structural biology. Notably, single-pass trans-membrane proteins, the topological classification of which PglC is predicted to belong to, are vastly under-represented in the PDB (Protein Data Bank) with only 0.5% (387 structures) of the non-redundant PDB having the same classification. Additionally, computational modeling of PglC structure using co-evolution and covariance constraints coupled with low sequence identity to proteins of known structure suggests the fold of the globular domain to be a novel protein fold (17).

1.6 Dissertation research specific aims

PglC, a putative single-pass transmembrane protein, represents the minimal catalytic unit for a structurally uncharacterized superfamily of PGT enzymes. The objective of my thesis research is to elucidate the unique structural characteristics of PglC that underpin membrane association and function of the minimal scaffold for the phosphoglycosyl transferase reaction in PGTs through work on the following specific aims:

Aim 1: Elucidate the structural determinants of membrane association.

Aim2: Identify the structural basis of enzymatic mechanism, and substrate

binding and specificity for PglC.

Aim3: Assess nanodiscs as a tool to study PglC in a native-like lipid environment by small angle x-ray scattering.

1.7 Summary of chapters in dissertation

The work presented in this dissertation explores the structural determinants of membrane association and enzymatic mechanism for PglC, an exemplar of the monotopic PGT superfamily. In the following chapter, the circuitous path required to determine the novel molecular structure of PglC from *C. concisus*, a related strain to *C. jejuni,* is discussed. Subsequently, chapters three and four, analyze the membrane interaction interface and structural determinants of the covalent-intermediate mechanism for PglC through structural, bioinformatics, and biochemical frameworks. Chapter Five lays the groundwork for understanding PglC, and therefore other small integral monotopic membrane proteins, in a native-like lipid in environment through the use of lipid-bilayer nanodiscs. Lastly, the major conclusions arising from the work presented in this dissertation are summarized in chapter six.

CHAPTER TWO

From crystal to structure of PglC – a circuitous structure solution

The work presented in this chapter is included in the following accepted article: Ray, L.C., Das, D., Entova, S., Lukose, V., Lynch, A.J., Imperiali, B., and Allen, K.N. Membrane association of monotopic phosphoglycosyl transferase underpins function. *Nature Chemical Biology* (2018). *(in press).*

2.1 Introduction

PglC, a monotopic PGT, is responsible for the first membrane-committed step of N-linked glycosylation in the pathogenic Gram-negative bacterium *C. jejuni*. To this end, PglC catalyzes the transfer of the unique sugar-phosphate bacillosamine from the UDP-activated form (UDP-diNAcBac) to a membrane-resident undecaprenol phosphate (UndP) (20). As the downstream glycoconjugates assembled through similar glycosylation pathways play roles in promoting pathogenicity and survival, targeting of PGTs through antibiotic intervention is well documented (6, 9, 29). However, due to the significant structural and mechanistic divergence of the two bacterial PGT superfamilies, the potent nucleoside antibiotics active against polytopic PGTs WecA and MraY, do not show inhibitory activity towards members of the monotopic PGT superfamily (20). Overall, development of novel antivirulence agents is of considerable interest and necessitates structural characterization of the monotopic PGT superfamily.

As discussed in the previous chapter, the small size (23 kDa, 200 aa) and its homology to the soluble domain conserved among the monotopic PGT subfamilies, makes *C. jejuni* PglC representative of the minimal catalytic unit for the monotopic PGT reaction, and therefore an optimal target for structure elucidation (17). However, the predicted topology of PglC, inclusive of a putative

single-pass hydrophobic TM domain, brings inherent challenges to structure determination. Although prevalent (30), single-pass TM proteins are immensely underrepresented in the PDB comprising only 0.5% of the non-redundant PDB, suggesting intrinsic, and often insurmountable, challenges in determination of their three-dimensional structures.

2.1.1 X-ray crystallography as a tool to elucidate structures of integral membrane proteins

Macromolecular x-ray crystallography is a remarkably powerful method for determining the three-dimensional structure of proteins to atomic resolution. In this method, a three-dimensional, highly-ordered array of proteins, in the form of a crystal, is required to produce diffraction patterns of sufficient quality and high resolution in order to resolve the atomic structure.

As production of high quality protein crystals is at the core of this structural biology method, use of crystallography to solve the structure of an integral membrane protein is fraught with challenges. First, membrane proteins are highly hydrophobic in nature and therefore poorly soluble in aqueous purification and crystallization buffers. Additionally, heterologous expression of these proteins is often challenging as a result of their frequent expression into inclusion bodies, or host-cell toxicity resultant of insertion of the integral membrane protein into host-

cell membranes (31). Lastly, to surmount the first two challenges, purification and crystallization of membrane proteins often requires the use of high concentrations of detergents to mimic the hydrophobic membrane environment, however the heterogeneous compositions of detergents can inhibit crystal growth (32). Even with considerable advances in membrane protein crystallography, many of the challenges inherent to studies with integral membrane proteins remain (33). A common strategy to surmount the challenges brought by the TM domains of integral membrane proteins is the "divide-and-conquer" approach to crystallography. This approach, utilizing natural domain boundaries within the protein to create truncation constructs with more amenable properties for purification and crystallization, often creates constructs omitting TM domains entirely. Frequently this leads to structure determination of the soluble domains of the protein, however the benefit must be weighed against the cost that any opportunity for structural insight into membrane association through the TM domain is therefore lost.

2.1.2 Prior optimization of crystallization constructs and conditions for PglC

Previous work towards the 3D-structure determination of PglC in the Allen and Imperiali groups adopted the "divide-and-conquer" strategy given that the

soluble domain, including the active site, of *C. jejuni* PglC is homologous to the soluble domains of the other monotopic PGT subfamilies (Figure 2.1). Contrary to expectation, the N-terminal hydrophobic helix was found to be required for robust expression, purification, and crystallization (34, 35). Additionally, this work identified PglC from an orthologous organism, *Campylobacter concisus*, with 72% sequence identity that lacks a C-terminal hydrophobic patch hypothesized to cause insolubility observed for *C. jejuni* PglC, to be more soluble and amenable to crystallography. A serine (S) – glycine (G) repeat linker (SGSG) was added between the N-terminus of PglC and the SUMO-purification tag to allow for efficient removal of SUMO (34). Initial crystallization conditions for wild-type (WT) *C. concisus* SGSG-PglC were determined through high-throughput screening (HTS) at the Hauptmann-Woodward Institute (HWI) crystallization facility. Optimization of the crystallization conditions to 300-600 mM $MgCl_2$, 100 mM Bis-Tris pH 6.0, and 25-30% polyethylene glycol 3350 in hanging-drop format produced crystals which produced high-quality diffraction patterns to a resolution of 3.00-3.38 Å (35). Although considered to be moderate, such diffraction limits should allow electron density for the Cα- back bone of the protein chain to be resolved.

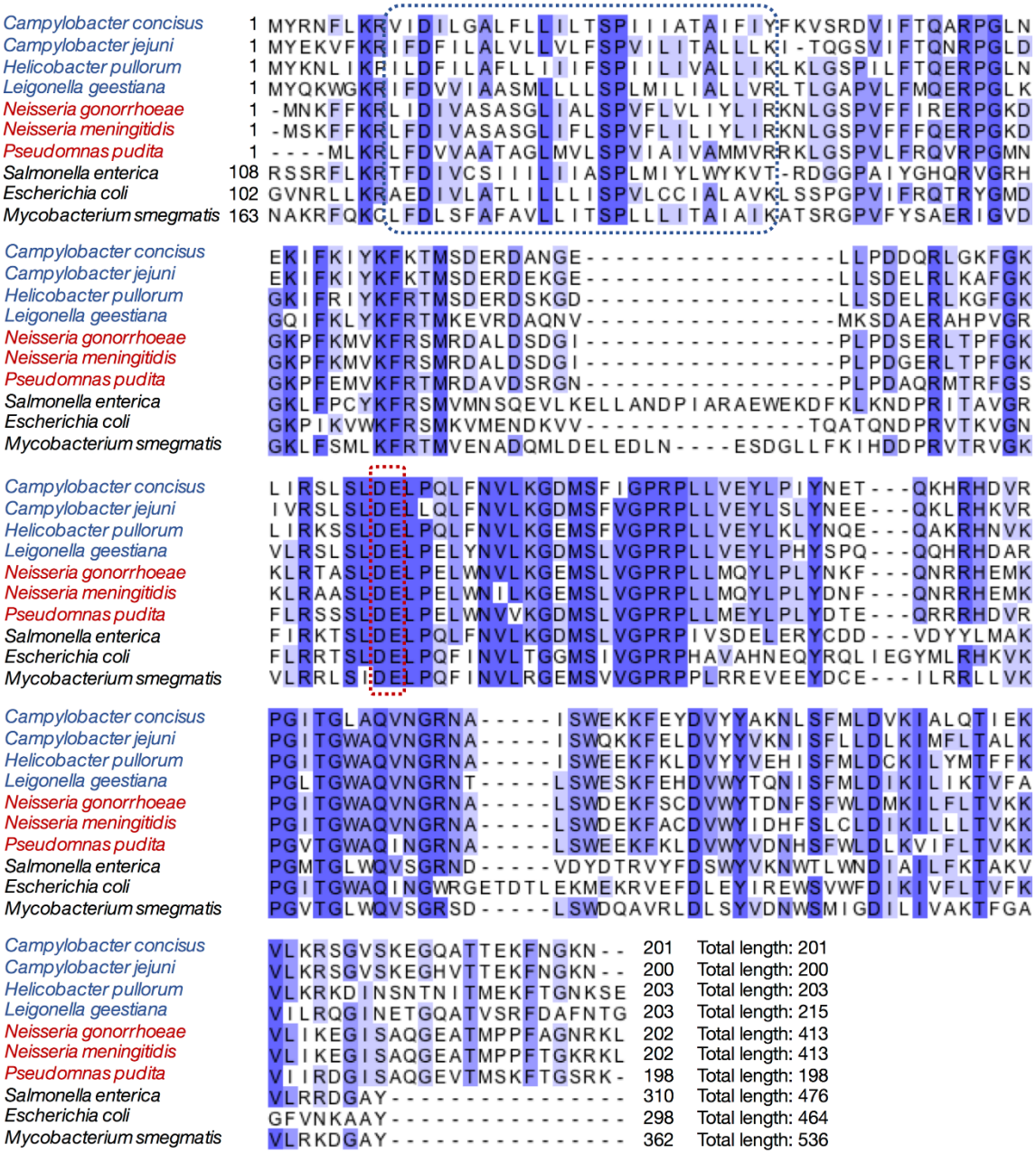

Figure 2.1. Sequence alignment of *C. jejuni* PglC-like minimal catalytic domain across all three monotopic-PGT superfamilies. Representative sequences from the PglC subfamily (BLUE): *C. concisus* PglC (A0A0M4SI81)/ *C. jejuni* PglC (O86156)/ *H. pullorum* PglC (E1B268)/ *L. geestiana* PglC (A0A0W0UA41). Representative sequences from the PglB-bifunctional subfamily (RED): *N. gonorrhoeae* PglB (A0A1D3HQ90)/ *N. meningitidis* PglB (Q9RR58)/ *P. putida* Sugar

transferase (A0A0P7CW64). Representative sequences from the WbaP subfamily (BLACK): *E. coli WcaJ (*P71241)/ *S. enterica* WbaP (S4IKQ0)/ *M. smegmatis* WcaJ (A0A0D6J209). The hydrophobic membrane associated domain is identified by a blue box. The active-site catalytic dyad characterized by Das, Kuzmic, and Imperiali (25) identified by the red box.

2.1.3 Overcoming the phase problem in x-ray crystallography

Although x-rays behave as waves with both amplitude and phase properties, in practice experimental detection is only able to record the intensity and angle of the diffracted ray while all accompanying phase information is lost. This is the "phase problem" inherent to macromolecular crystallography (36). As a result, the atomic model of the protein structure can only be built when patterns of recorded diffraction intensities are recombined via Fourier transformations with calculated phases obtained by a phasing method into an interpretable electron density map. The phases required for structure elucidation are primarily obtained via molecular replacement (MR) or an experimental phasing method, or a combination thereof.

Obtaining phase information for the unknown structure via MR is increasingly common as new macromolecular structures are added to the PDB routinely. MR is an algorithmic method, which relies and rotation and translation functions to correctly place the atomic coordinates of a known structure, serving

as a MR search model, in the unknown unit cell (37). This ultimately requires a high degree of similarity between the known and unknown structures. In general, for MR to be successful in providing phases for an unknown structure, the search model and unknown structures must share approximately 40% sequence identity and cover approximately 50% of the structure (38). As PglC has low sequence similarity and homology to proteins of known structure in the PDB, previous attempts to phase datasets of WT SGSG-PglC used computationally derived models as for MR search (17, 35). Even though the computational models created by EV-fold and I-TASSER are produced through employment of evolutionary sequence constraints to inform possible tertiary structure conformations, they were ineffective as MR search models for the native datasets of WT-PglC.

For proteins like PglC which lack MR search models with sufficient structural homology to produce viable phases, experimental phasing methods are required. One of the many methods developed to determine phases experimentally, exploits the anomalous dispersion from atoms heavier than the H, C, N, and O atoms that typically comprise most proteins (36, 39). This dispersion arises from the capacity of heavy atoms to absorb a fraction of the incident x-rays at unique wavelengths. Consequently, Friedel's law is effectively broken as the reflections $hkl \neq \overline{hkl}$ in intensity. The absorption of x-rays from atoms commonly

found within proteins is negligible at the wavelengths used for crystallographic experiments. Therefore, any differences in intensity observed in the diffraction patterns collected correspond to absorption by the heavy atoms within the structure.

In theory, single-wavelength anomalous dispersion (SAD) phasing is the simplest case of phasing via the exploitation of anomalous dispersion. SAD is accomplished through the collection of diffraction data at a single wavelength corresponding to the elemental absorption edge which elicits the most anomalous scattering from the heavy atom, and therefore a peak in the f' spectra (36). The anomalous scattering coefficient, f'', provides a measure of the peak anomalous scattering power for the element. The heavy atoms central to SAD experiments can be endogenous to the protein or incorporated through a variety of methods (40). Anomalous scattering from endogenous, native sulfur atoms present in cysteines and methionines within the protein sequence can be measured at longer wavelengths (>1.54 Å or 8 keV) where f'' is approximately equal to 0.57 scattering electrons per S atom (Figure 2.2). Although it is theoretically possible to measure anomalous scattering, and therefore phases using sulfur-SAD (S-SAD), the low anomalous signal from S is often insurmountable (41). As a result, it significantly more common to utilize SAD methods in which the heavy atoms have been added

to the protein sample. Of these, the most common method of incorporation of heavy atoms is through the substitution of selenomethionine (SeMet) for methionine in the primary sequence of the protein during protein expression (42). Thereby, replacing the low anomalous signal S atoms with selenium. Substitution of SeMet for methionine residues allows for significant anomalous scattering power upon data collection at ~1 Å (12.6 keV) with f'' = 3.77 e^- (Figure 2.2).

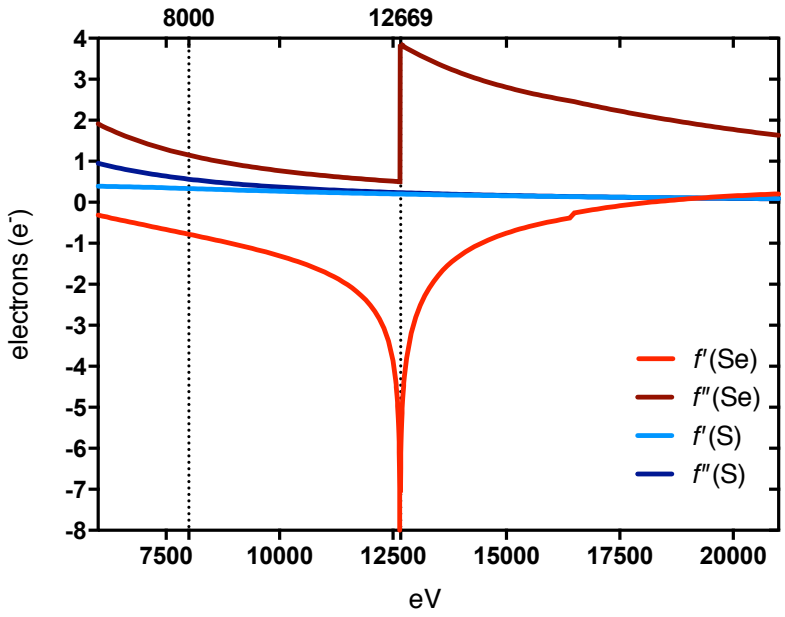

Figure 2.2. Theoretical f′ and f″ plots for sulfur and selenium. S-SAD datasets collected 8000 eV result in the theoretical absorption of 0.57 e^- per S atom. Se-SAD datasets collected at the peak energy (12669 eV) result in the theoretical absorption of 3.77 e^- per Se atom. Plot range encompasses energies accessible by synchrotron x-ray source beamline 24-ID-C (NE-CAT) at Advanced Photon Source (APS) in Chicago, Il (6 keV – 21 keV). Plot created with data tabulated by http://skuld.bmsc.washington.edu/scatter/AS_periodic.html.

2.1.4 Towards the structure of PglC

The work presented in this chapter summaries the path from crystal to the structure of *C. concisus* PglC, the first example of the monotopic PGT superfamily. First, this chapter discusses the optimization of crystallization conditions for the three variants of *C. concisus* PglC ultimately required to solve the structure. Next, the attempts and eventual success of determining experimental phases for PglC are explored through the use of SAD phasing to calculate initial interpretable electron-density maps, which were further enhanced through exploitation of non-crystallographic symmetry (NCS) within the assymetric unit (ASU), and iterative model building. The subsequent use of MR to phase higher-resolution data-sets of PglC leading to the final model is discussed. Ultimately, the final refined structure of full-lengthPglC encompassing both soluble and membrane-associated domains determined via SAD phasing and subsequent MR using wild-type SeMet-substituted protein and I57M/I87M and I57M/Q175M variants is revealed. Comparison to extant folds confirmed that PglC comprises a novel protein fold. Lastly, the overall agreement of the final model for PglC with computationally-derived models is discussed. The novel structure of PglC presented in this chapter not only enables homology modeling for the entire bacterial monotopic PGT

superfamily, but also adds an example of a monotopic protein complete with its membrane-associated domain to the PDB. Thus, the molecular knowledge gleaned from PglC in the subsequent chapters can be translated across the ~15,000 members of the monotopic PGT superfamily, and monotopic proteins overall.

2.2 Materials and Methods

2.2.1 Crystallization of PglC constructs for experimental phasing

Preparation of the three variants (WT-SeMet, I57M/I87M, and I57M/Q175M) of *C. concisus* used for crystallography was performed by Dr. Debasis Das (Imperiali Lab, MIT) and provided at a concentration of 260-276 μM in 50 mM HEPES pH 7.5, 100 mM NaCl, and 0.03% DDM. Notably, the constructs of PglC co-purified with approximately 3 phosphatidyl ethanolamine : 1 phosphatidtyl glycerol endogenous lipid observed by thin layer chromatography. Final crystallization conditions of 0.1 M Bis-Tris pH 6.0, 0.4 M $MgCl_2$, and 23-27% PEG 3350 were optimized for a protein concentration of 260-276 μM for the three constructs of PglC. Additional detergent was not added during crystallization. Hanging-drop crystallization experiments were set up at 17 °C with solutions purchased from Hampton Research. All reagents were cooled to 17 °C prior to set

up of crystallization experiments. I57M/I87M PglC at a concentration 260 μM was co-crystallized with 260 μM UndP (using a 10 mM stock solution of UndP in DMSO) following a 30-minute incubation step on ice. I57M/I87M PglC crystals used for data collection appeared within 7 days. I57M/Q175M PglC crystals were grown under similar conditions (0.1 M Bis-Tris pH 6.0, 0.4 M $MgCl_2$, 23% PEG 3350) at 17 °C with temperature equilibrated solutions. Co-crystallization experiments with 5-iodo-UDP were carried out following incubation of 1 mM 5-iodo-UDP in 260 μM protein on ice for 30 minutes. I57M/Q175M PglC crystals used for data collection appeared within 7 days. WT Se-Met PglC crystals were grown at 17 °C with temperature equilibrated crystallization conditions of 0.1 M Bis-Tris pH 6.0, 0.3 M $MgCl_2$, 27% PEG 3350, and 1 mM TCEP. Se-Met PglC at a concentration of 276 μM was co-crystallized with 1 mM UDP after incubation on ice for 30 minutes. WT Se-Met PglC crystals used for data collection appeared within 3 days and reached their final size after 14 days. Addition of cryoprotectants resulted in decrease in diffraction limits for all crystals tested, thus crystals used for data collection were not subjected to cryoprotection. All crystals used for data collection were flash-cooled by plunging into liquid nitrogen for transport and data collection.

Table 2.1. Amino acid sequences of constructs used in structure solution of PglC. Underline denotes N-terminal methionine residue. Yellow represents SeMet. Red illustrates position of the point mutations.

Construct	Amino acid sequence
WT SeMet	SGSGMYRNFLKRVIDILGALFLLILTSPIIIATAIFIYFKVSRDVIFTQA RPGLNEKIFKIYKFKTMSDERDANGELLPDDQRLGKFGKLIRSLSLDELP QLFNVLKGDMSFIGPRPLLVEYLPIYNETQKHRHDVRPGITGLAQVNGRN AISWEKKFEYDVYYAKNLSFMLDVKIALQTIEKVLKRSGVSKEGQATTEK FNGKN
I57M/Q175M	SGSGMYRNFLKRVIDILGALFLLILTSPIIIATAIFIYFKVSRDVIFTQA RPGLNEKIFKMYKFKTMSDERDANGELLPDDQRLGKFGKLIRSLSLDELP QLFNVLKGDMSFIGPRPLLVEYLPIYNETQKHRHDVRPGITGLAQVNGRN AISWEKKFEYDVYYAKNLSFMLDVKIALMTIEKVLKRSGVSKEGQATTEK FNGKN
I57M/I87M	SGSGMYRNFLKRVIDILGALFLLILTSPIIIATAIFIYFKVSRDVIFTQA RPGLNEKIFKMYKFKTMSDERDANGELLPDDQRLGKFGKLMRSLSLDELP QLFNVLKGDMSFIGPRPLLVEYLPIYNETQKHRHDVRPGITGLAQVNGRN AISWEKKFEYDVYYAKNLSFMLDVKIALQTIEKVLKRSGVSKEGQATTEK FNGKN

2.2.2 *Data collection and processing*

A WT Se-Met PglC dataset collected from a crystal diffracting to 3.11 Å at beamline 24-ID-C at the Advanced Photon Source (Chicago, IL) at the Se X-ray absorption energy peak (12665 eV) allowed initial partial phases to be solved by SAD using the Phenix suite (43). Matthews coefficient analyses for the data set with increased unit-cell dimensions (a = b = 143. 375 Å, c = 194.004 Å; P 3_1 2 1) suggested 8 copies in the ASU. WT Se-Met data were scaled and integrated using XDS (44). SHELXD (45) was run for 5000 trials with a resolution cut-off of 4.5 Å to identify 16 Se sites. Phenix.SOLVE (46) was used to find an additional 6 Se sites

and calculate subsequent Se substructure phases for 22 out of the expected 32 Se atoms in the ASU. Phenix.RESOLVE (47) was used to perform initial solvent flattening and phase-extension.

2.2.3 Model building

Alpha-helical density apparent in the solvent-flattened map allowed for initial building of poly-Ala helical fragments manually. Using these α-helical fragments as a starting model, Phenix.AutoBuild (48) was able to locate and assign sequence to two copies of the RMH in the ASU. Phenix.Find_NCS was used to find the two-fold NCS operators relating the two helices, and the two-fold used for NCS map-averaging in Phenix.NCS_average. Phenix.Find_NCS was used to find all 8 NCS related positions from the electron density in the two-fold averaged map. The two-fold averaged map was further averaged over all eight NCS operators by Phenix.NCS_average. Additional residues were built manually into the density with sequence assigned, using as a guide computationally-derived models of PglC from EV-fold (*Campylobacter jejuni*) and RaptorX (49), as well as the model of *E. coli* WcaJ (Pfam accession P71241) computed by covariance using Rosetta (50). This manually-extended model containing a dimer of 94 residues (AAs 3-60, 74-98, 165-175) in each chain was used as a search model for molecular replacement via

Phenix.PhaserMR (51) into a native I57M/I87M 2.59 Å dataset collected at beamline 24-ID-C at the Advanced Photon Source (Chicago, IL) containing 2 copies in the ASU with smaller unit cell dimensions (a = b = 70.802 Å, c = 188.442 Å; P 3_2 2 1). Phenix.AutoBuild (48) was used to complete building of 86% of the model and an additional 14 residues were built manually into the electron density using COOT (52). This model containing two subunits with 185 residues in each chain was used to phase a more complete, higher I/σ(I) dataset of I57M/Q175M PglC at 2.74 Å resolution. The overall fold of the WT and variant structures are identical despite crystallizing in enantiomeric space groups of P 3_1 2 1 for the native datasets and P 3_2 2 1 for the SeMet-derivatized dataset. The space-group change is a result of subtle rearrangements as the crystal contacts within the lattice are unchanged between space groups. .

2.2.4 Model Refinement

Refinement against the electron density map was performed with Phenix.Refine(53) to refine XYZ coordinates, real-space, rigid body, and group B-factors. Subsequent rounds of refinement included refinement of translation-libration-screw (TLS) parameters, manually placed waters, and simulated annealing of Cartesian coordinates and torsion angles. The final model with two

subunits in the asymmetric unit was refined to R_{work}/R_{free} of 0.2587/0.2815 with no significant outliers using Phenix.Refine (53). Both chains of the model contain 185 out of 205 amino acids. Chain A contains amino acids (-)3 to 182 and chain B contains amino acids (-)2 to 183 (the minus signs indicate non-native residues resulting from addition of expression tags). Chains A and B are highly similar with an RMSD of 0.31 Å. The extended loop structure encompassing residues 62-81 is well ordered in chain B owing to its participation in crystal contacts, however only weak density for this loop was observed in chain A. Residues 148- 153 were not well resolved in either chain and were placed into $2F_o$-F_c density contoured at 1 RMSD in COOT(52). The final model of PglC was refined with two protein chains, four molecules of phosphatidylethanolamine (PE), two Mg^{2+} ions, one inorganic phosphate ion, and one PEG (tetraethylene glycol). The chemical structures of endogenous, natural abundance PE were built into the model using electron density maps calculated with the coefficients $2F_o$-F_c and F_o-F_c resulting in 4 molecules with differing acyl chain lengths (two PE per monomer of PglC). Molecules 303 (Chain A) and 302 (Chain B) lie within the predicted membrane interaction surface for each monomer of PglC, while molecules 304 (Chain A) and 303 (Chain B) are not believed to be in physiologically relevant positions. Exclusion of the PE molecules with non-physiological positions increased both

R_{work} and R_{free} statistics during refinement. Additionally, diffuse scattering from disordered detergent and additional unobserved lipid molecules could have contributed to slightly elevated R_{work}/R_{free} values, as extensive analysis of Matthews coefficients and examination of special projections excluded twinning as a factor. An inorganic phosphate ($PO4^{-3}$) ion and a Mg^{2+} ion were modeled in Chain A, and a Mg^{2+} ion and ordered water molecule were modelled in Chain B into positive F_o-F_c density contoured to 4 RMSD in COOT (52). The PEG molecule was modeled into positive F_o- F_c density in Chain A contoured to 2.5 RMSD in COOT (52).

2.3 Results and Discussion

2.3.1 Crystallization of PglC Constructs for Experimental Phasing

We were unable to obtain sufficient phase information using the published covariation model (17) for structure solution via MR from datasets collected by Andrew Lynch from WT crystals. Thus, three variants of SGSG-PglC were prepared by Debasis Das (Imperiali Lab, MIT) to pursue structure solution via experimental phasing techniques: I57M/I87M, I57M/Q175M and selenomethionine (Se-Met) derivatized WT (Table 2.1). The I57M/I87M and

I57M/Q175M variants were engineered with additional methionines to allow for increased signal for phasing via native sulfur-SAD. Crystallization of both variants was obtained through optimization of the final crystallization conditions for WT *C. concisus* SGSG-PglC identified by Andrew Lynch. Final crystallization conditions for I57M/I87M PglC co-crystallized with UndP (0.1 M Bis-Tris pH 6.0, 0.4 M $MgCl_2$, 23% PEG 3350; 2 μL PglC: 1.5 μL well solution) produced crystals which diffracted to 2.59 - 3.0 Å (Figure 2.3). Crystals of I57M/Q175M PglC were optimized in similar conditions (0.1 M Bis-Tris pH 6.0, 0.5 M $MgCl_2$, 25% PEG 3350; 2 uL PglC: 2 uL well solution) to produce crystals which diffracted to 2.74 - 3.5 Å (Figure 2.3). Indexing of the datasets collected for both methionine variants found unit-cell dimensions of a = b = 70.802 Å, c = 188.442 Å for I57M/Q175M, and a = b = 71.61 Å, c = 189.442 Å for I57M/I87M. Matthews coefficient analyses for these data sets suggests 2 copies in the asymmetric unit (ASU). In addition, WT SeMet derivatized PglC was crystallized by optimization of the initial crystallization condition to a final crystallization condition of 0.1 M Bis-Tris pH 6.0, 0.3 M $MgCl_2$, 27% PEG 3350, 1 mM TCEP, and 1 mM UDP. Optimized WT Se-Met crystals diffracted to between 3.0 – 3.5 Å resolution (Figure 2.3). Notably, derivatization with Se-Met resulted in doubling of unit cell axes a and b to give a larger unit cell with dimensions a = b = 143.375 Å, c = 194.004 Å. As a result of the doubling of the

a and b unit cell axes, Matthews coefficient analysis suggests ASU composition increased from 2 subunits to 8 subunits of PglC.

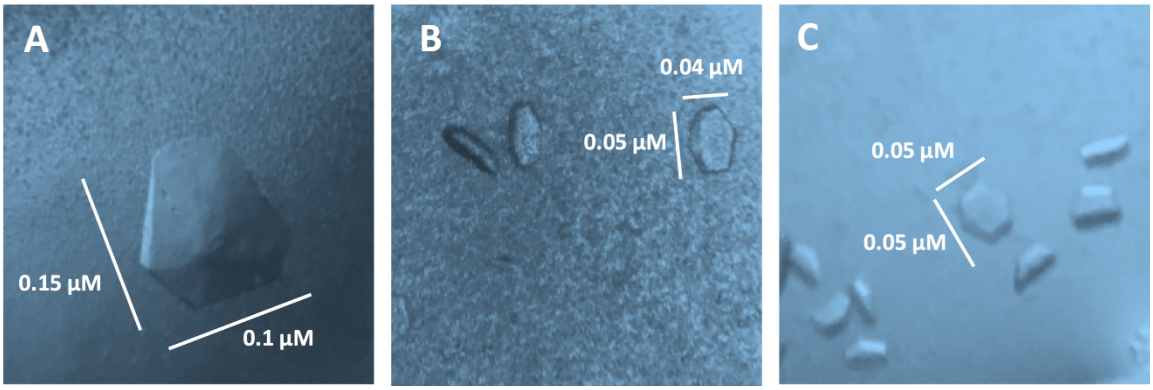

Figure 2.3. Representative crystals of *C. concisus* PglC constructs used for experimental phasing. A, WT-SeMet. B, I57M/Q175M. C, I57M/I87M PglC.

2.3.2 *Experimental phasing of PglC datasets*

Pursuing experimental phase information from native S-SAD is challenging and requires high data quality and exceedingly high redundancy to enrich the native sulfur anomalous signal (f″ = 0.57 e^- at collection wavelength of 1.5498 Å) (41). Despite using the engineered PglC constructs with additional methionines to increase the sulfur signal (I57M/I87M and I57M/Q175M), initial datasets of PglC did not contain sufficient anomalous signal for phasing via SAD. For the data sets collected, the anomalous measurability, a measure of the fraction of Bijvoet related intensity differences where $|\Delta I|/\Delta\sigma I > 0.3$, observed for these data sets remained

below the threshold for sufficient signal (> 0.05) at useful resolution ranges. A second measure of anomalous signal strength (CC_{anom}) also remained below the usable threshold for all datasets indicating insufficient signal to obtain phases by S-SAD (54). Ultimately, the moderate diffraction (≥ 2.74 Å) of the two methionine variant construct crystals, did not provide sufficient data quantity and quality for successful native S-SAD phasing.

To circumvent the low signal-to-noise inherent in native S-SAD data for PglC, the Se-Met derivatized WT-PglC construct was utilized to increase the anomalous phasing power. Selenium atoms have approximately 7.5-times more anomalous scattering power with f''= 3.77 e^- at a data-collection wavelength of 0.9792 Å compared to 0.56 e^- scattered by sulfur. From the Se-Met PglC crystals, a dataset from Se-Met WT PglC was collected to a resolution of 3.11 Å in the large unit cell. Initial phasing of this dataset was accomplished with the assistance of Dr. Raj Rajankrishnar (APS, NE-CAT). The selenium substructure was partially solved by SHELXD (45) which found 16 of 32 Se-sites expected for the large unit cell (8 Copies * 4 Se/ PglC). Notably, a suitable solution was only found using an exhaustive search of 5,000 trials at a low-resolution threshold of 4.5 Å; fewer trials and higher-resolution cut-offs did not yield viable heavy-atom substructure solutions (Figure 2.4). To quantify the plausibility of solutions found by SHELXD

(45) the correlation coefficient (CC) between the observed structure factors and calculated structure factors between all of the reflections (CC_{all}) and reflections not used in calculation (CC_{weak}) are compared. A good indication of correct heavy atom substructure solutions calculated by SHELXD is separation of some solutions over the bulk of the solutions and a high CC_{all} versus $CC_{weak} \geq 30\%$ (45). The low-resolution of the measurable anomalous signal for this dataset which dropped below 0.05 at 4.0 Å contributes to the necessity of a low-resolution cut-off and large number of trials for substructure search. Using the initial sites found by SHELXD (45), the selenium substructure was extended to 22 sites with Phaser (51) and subsequent determination of phases allowed for the calculation of a weakly interpretable electron density map allowing the placement of polyAla helices into the electron density.

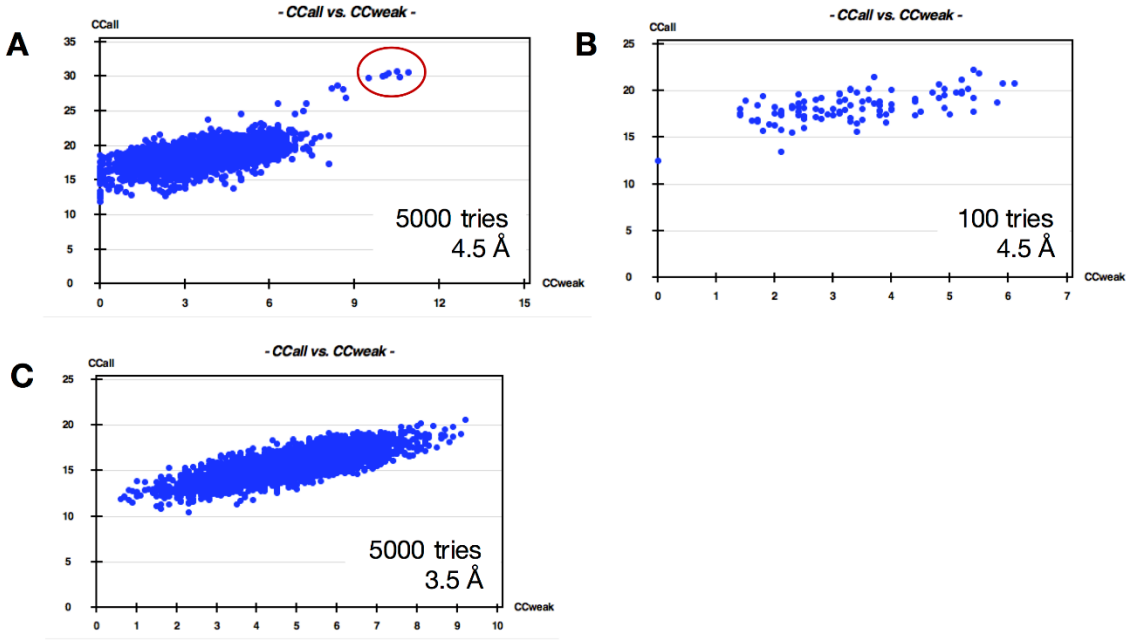

Figure 2.4. Correlation coefficient plots indicating quality of substructure solutions as output from SHELXD during Se-atom substructure search visualized through HKL2MAP (55). **A,** CC_{all} vs. CC_{weak} for 5,000 SHELXD solutions calculated at 4.5 Å resolution. Plausible solutions with $CC_{al} \geq 30\%$ and separation over the bulk solutions circled in red. **B,** CC_{all} vs. CC_{weak} for 100 SHELXD solutions calculated at 4.5 Å resolution. **C,** CC_{all} vs. CC_{weak} for 5000 SHELXD solutions calculated at 3.5 Å resolution, the resolution maximum for anomalous signal calculated by SHELXC (56).

2.3.3 Building the final model of C. concisus PglC

Initial phases and maps were improved through a combination of AutoBuild, NCS averaging, manual iterative model building guided by computational models to yield a model with sequence placed for 46% (residues 3-60, 74-90) of PglC (Figure 2.5). An additional dataset of I57M/I87M to higher

resolution in the small unit cell (a = b = 71.61 Å, c = 189.442Å) was collected and the model from the WT SeMet dataset was used as a molecular replacement search model for the higher resolution (I57M/Q175M) dataset (2.59 Å). Molecular replacement successfully placed two protomers in the ASU allowing for AutoBuild to extend the model with an additional 32 residues. Notably, molecular replacement was only successful in the enantiomeric space group of P 3_2 2 1. An origin shift of approximately 0.5 between the small cell (Met Mutants) and the large cell (SeMet) datasets explains the change of space group from P 3_1 2 1 for the small cell to P 3_2 2 1 for the large cell. Notably, attempts to use the SeMet-dataset-derived model as a molecular replacement search model for the 2.74 Å dataset did not produce a viable solution. From this result, it is clear that the increase in resolution in 2.59 Å dataset, though minimal, provided critical data for molecule placement resulting in an acceptable solution. From the MR solution in the 2.59 Å dataset, the remaining 20 residues were built manually into the density to complete the model of PglC. Due to low data quality, in particular low signal-to-noise, in the highest resolution shells for the 2.59 Å dataset, isomorphous replacement was used with the model encompassing residues 3-60, 74-140, and 165-180 to phase the 2.74 Å dataset. Ultimately, the final model of PglC was determined via single-wavelength anomalous dispersion phasing using wild-type

selenomethionine-substituted protein and I57M/I87M and I57M/Q175M variants. The model of full-length I57M/Q175M PglC from *C. concisus*, encompassing both soluble and membrane-associated domains was fully refined with excellent geometry to an R_{work}/R_{free} = 0.2587/0.2815 (Figure 2.6; Table 2.2).

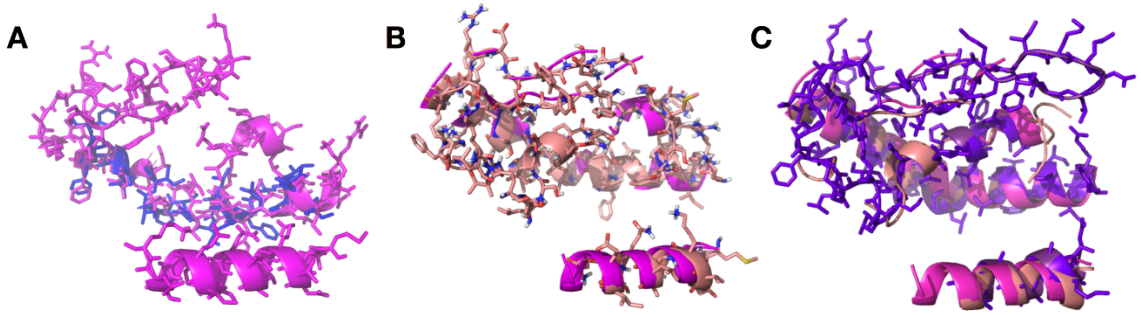

Figure 2.5. Progress of iteratively built model of PglC. A, Computational-model guided placement of main-chain and subsequent placement of side-chains (pink) allowed extension from the initial polyAla helix (blue). B, 2-fold NCS averaging of the electron-density map enabled extension of the initial model (pink) to the extended model (salmon). C, 8-fold NCS averaging of the electron-density map enabled extension (from salmon to purple) of the model with 46% coverage of the *C. concisus* PglC sequence (purple).

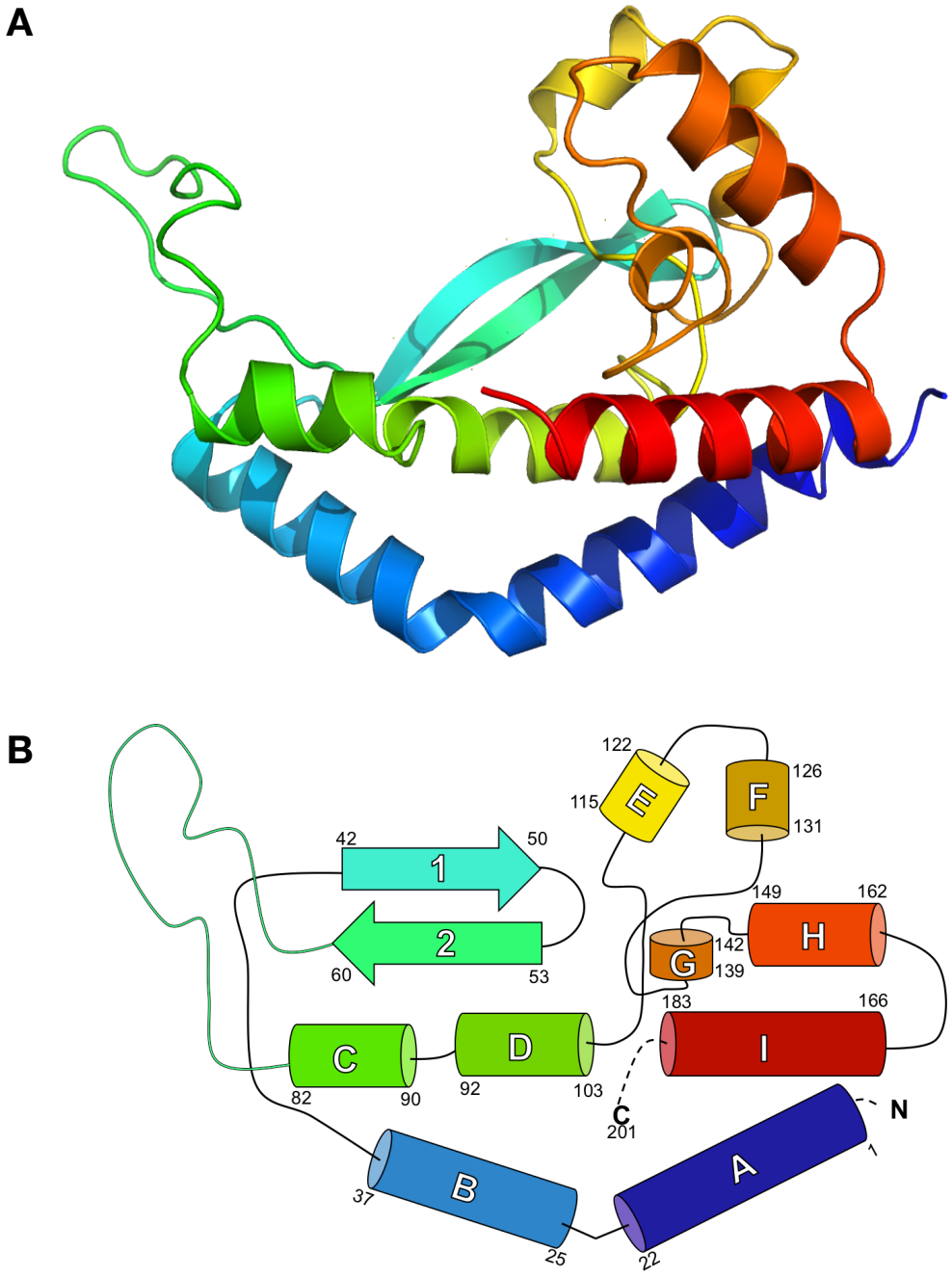

Figure 2.6. The final refined model of *C. concisus* I57M/Q175M SGSG-PglC. A, Cartoon representation of structure colored from N-terminus (blue) to C-terminus (red). **B,** Topology diagram for the novel fold of PglC.

Table 2.2. Data collection and refinement statistics.

	I57M/Q175M Variant	I57M/I87M Variant	WT Se-Met
Data Collection			
Beamline	APS 24-ID-C	APS 24-ID-C	APS 24-ID-C
Wavelength (Å)	1.5498	0.9792	0.9791
Resolution range (Å)	62. 81 - 2.74 (2.84-2.74)	94.93 - 2.59 (2.71 - 2.59)	124.17- 3.11 (3.22-3.11)
Space group	P 3_2 2 1	P 3_2 2 1	P 3_1 2 1
Unit Cell (Å)	a = b = 70.802 c = 188.442	a = b = 71.61 c = 189.442	a = b = 143.375 c = 194.004
Total Reflections	272497 (22932)	70621 (8076)	204661(77521)
Unique reflections	27735 (1941)	18030 (2131)	41202 (1010)
Multiplicity	18.0 (11.8)	3.8 (3.8)	5.0 (3.4)
Completeness	0.99 (1.0)	0.98 (0.97)	0.97 (0.97)
Mean I/sigma(I)	21.4 (2.5)	11.6 (0.7)	10.5 (0.9)
Wilson B-factor	68.21	70.69	75.42
R_{merge}	0.0987 (1.1)	0.084 (1.708)	0.118 (1.414)
R_{meas}	0.1016 (1.159)	0.098 (1.981)	0.148 (1.839)
$CC_{1/2}$	1.0 (0.71)	0.99 (0.321)	0.99 (0.36)
CC*	1.0 (0.893)		
Refinement			
Ref. used in refinement	27735 (1941)		
Reflections used for R_{free}	1403 (127)		
R_{work}	0.2587 (0.3334)		
R_{free}	0.2815 (0.3663)		
CC_{work}	0.801(0.673)		
CC_{free}	0.879 (0.603)		
Number of non-H atoms	3145		
macromolecules	3043		
ligands	82		
solvent	20		
Protein residues	370		
RMS(bonds)	0.003		
RMS(angles)	0.66		
Ramachandran favored	97.54 %		
Ramachandran allowed	2.19 %		
Ramachandran outliers	0.27%		
Rotamer outliers (%)	0		
Clashscore	6.52		
Average B-factor	79.75		
macromolecules	79.49		
ligands	93.44		
solvent	64.34		

2.3.4 Molecular Structure of PglC adopts a novel fold

The final model of the catalytically-active I57M/Q175M variant will be designated as PglC throughout this dissertation. As discussed previously, PglC crystallizes with two protomers in the asymmetric unit. However, previous characterization of PglC in lipid bilayer nanodiscs via gel densitometry supports a functional monomeric biological assembly (57). The structure of PglC, comprising residues 1-183 (of the total 201, with observation of a GSG linker at the N-terminus), showed clear electron density (Figure 2.7) for the majority of the structure.

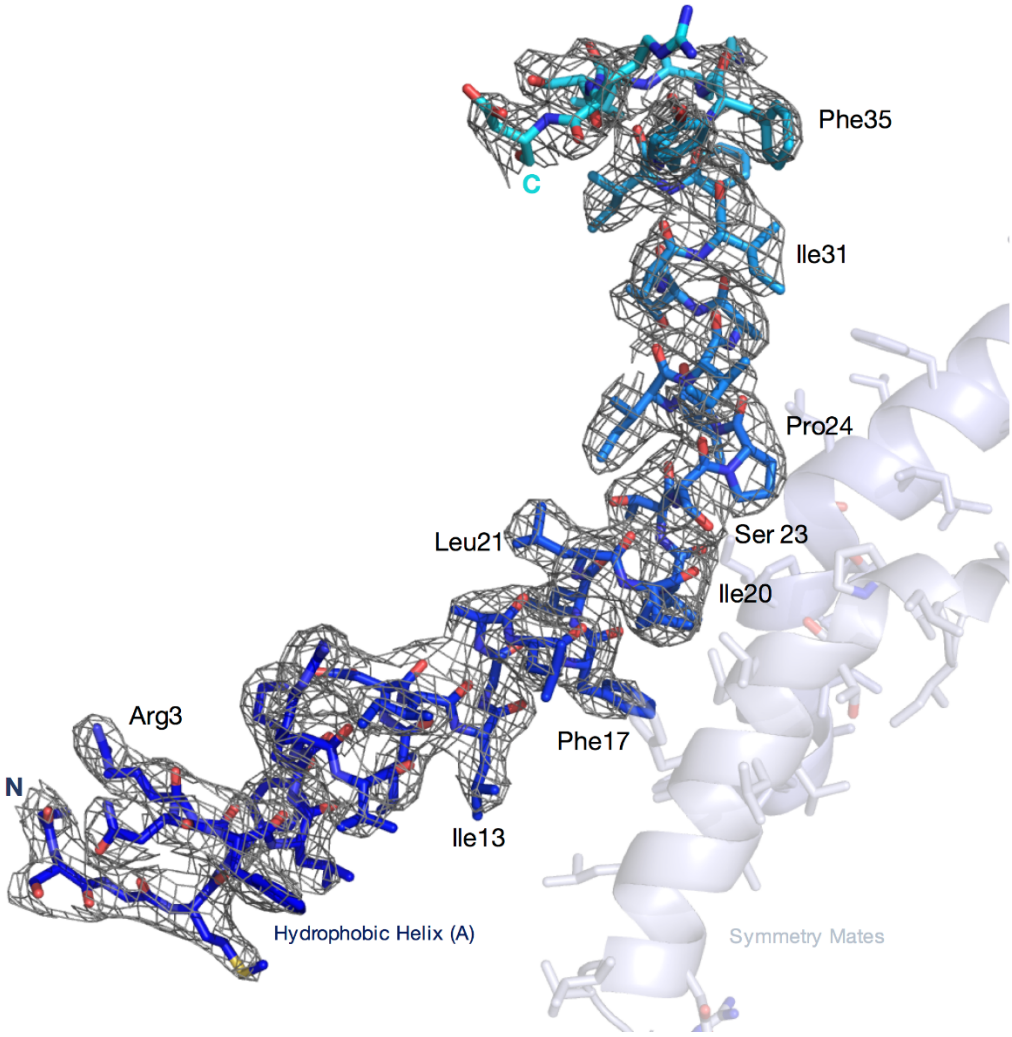

Figure 2.7. Representative electron density observed for PglC. Final $2F_o$-F_c electron density (gray mesh) contoured to 2σ in the region of the hydrophobic N-terminal helix (blue). Crystallographic symmetry mates (blue-gray) illustrating interdigitating Pro residues from the hydrophobic helices.

Overall, the structure of PglC reveals a new protein fold (Figure 2.6), where the extramembrane surface comprises a unique α-helix-associated β-hairpin motif, designated as the AHABh-motif (strands 1 and 2 and helix D; Figure 2.6B). Overall,

the structure is α-helical with the exceptions of the long β-hairpin of the AHABh-motif, an extended loop joining the β-hairpin to the first two amphipathic helices (helices C and D; Figure 2.6B), and a double twisted loop (residues 105-140) contributing to the only globular domain of PglC. From the N-terminus, α-helices A and B form a helix-break-helix motif which extends into a long β-hairpin structure (strands 1 and 2). An extended loop structure links the β-hairpin motif to planar helices C and D to create the AHABh-motif. The C-terminal end of helix D supports the base of the globular double twisted loop domain formed by the loops connecting helices E, F, G, and H. Lastly, helix I, co-planar with helices C and D, encompasses the C-terminal end of the resolved structure (Figure 2.6B).

It is important to note that crystal contacts resulting in the formation of the crystallographic dimer place the extended loop of one protomer impinging upon the space between the AHABh-motif β-hairpin and extended loop of the other protomer (Figure 2.8A). Although these crystal contacts are near the active site, the steric volume of the impinging loop does not occlude the active-site, and thus does not preclude further analysis of the structure of the active site. Each protomer of the crystallographic dimer contributes to a tetramer of symmetry mates to build the crystal lattice (Figure 2.8B). In each tetramer, the PglC protomer is arranged with the hydrophobic helix facing inwards towards the center with the proline

(Pro24) residue forming the kink interdigitating with corresponding proline (Pro24) on the proximal protomer. The packing of PglC within the crystal lattices explains the necessity for inclusion of the hydrophobic helix for crystallization of PglC to occur as previously observed (35). Furthermore, as described below, the hydrophobic helix is integral to the protein fold, and does not comprise a separate membrane embedded domain.

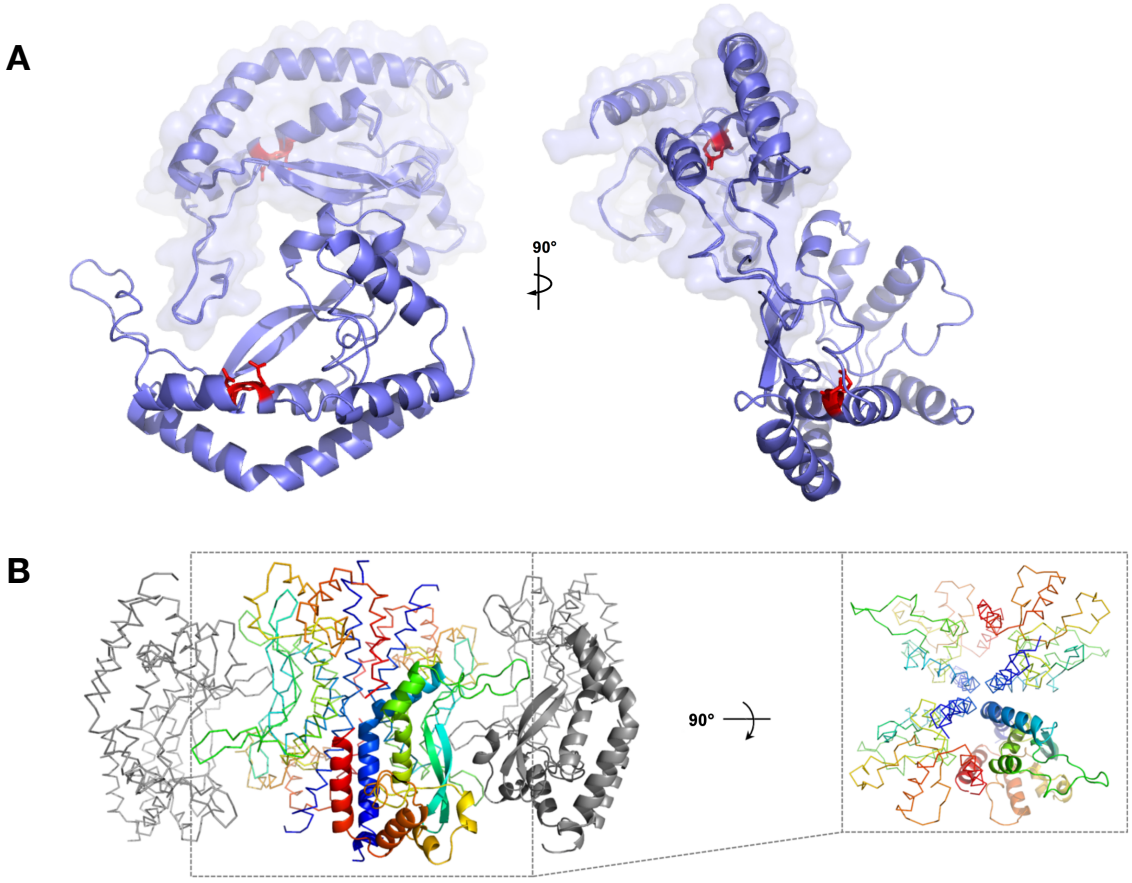

Figure 2.8. Crystal packing observed for *C. concisus* I57M/Q175M SGSG-PglC. **A,** Crystal contacts near the active site do not preclude analysis of the PglC active site. Crystallographic dimer of PglC shown in cartoon representation. Molecular surface of Chain B displayed to illustrate solvent excluded volume of the monomer. Catalytic Asp-Glu dyad shown in red sticks. **B,** Crystal contacts are mediated through the RMH neighboring symmetry mates. Crystallographic ASU depicted in cartoon and symmetry mates are depicted in ribbon representations.

2.3.5 Confirmation of novel protein fold

As described previously, the structure of PglC adopts a novel fold characterized by the AHABh-motif and kinked, hydrophobic helix. This was expected due to the low sequence identity to proteins of known structures in the PDB, as well as the failure of computationally derived models from EV-fold and I-TASSER for phasing via molecular replacement (17, 35). To confirm that the fold of PglC is novel, Chain A of the PglC model was used as the query structure for three-dimensional protein fold comparison to known structures in the PDB by the DALI protein structure comparison server (58). As expected, no significant tertiary structure matches to PglC were reported by DALI. Furthermore, the only secondary structure elements showing structural similarity were matches to sections of PglC helices C and D by long alpha-helices from a large variety of proteins from designed proteins, viral envelope proteins, and components of nuclear pore complexes from fungi (Table 2.3).

Table 2.3. Top 10 matches from DALI (58) search with final PglC model as query

Rank	PDB - Chain ID	Z-score	RMSD (Å)	# of aligned residues	# of aa in matched structure	% ID	Molecule Description
1	5j0i-A	4.3	3.7	61	73	8	Designed protein 2l6hc3_12
2	3csx-B	4.0	3.3	61	71	7	Putative uncharacterized prot.
3	5j2l-A	3.9	4.4	60	76	2	Protein design 2l4hc2_11
4	4knh-B	3.8	8.8	90	857	11	Nup192p
5	4knh-A	3.8	8.8	90	914	11	Nup192p
6	3hr0-B	3.8	8.2	74	249	7	Cog4
7	4i0x-E	3.8	5.1	61	67	3	Esat-6-like protein mab_3112
8	1jq0-A	3.7	4.4	57	71	2	Gp41 envelope protein
9	2vs0-B	3.7	6.5	68	84	9	Virulence factor esxa
10	4i0x-K	3.7	5.6	65	77	3	Esat-6-like protein mab_3112

2.3.6 Agreement of final model of PglC with computationally-derived models

In determining the structure of PglC, three computational models were utilized to guide iterative model building in weak electron density maps. In order to debias placement of PglC main chain within the density, their placement was guided by an overlay of the three models computed from independent sources. Two of the models, the EV-fold (17) and Rosetta (50) derived models utilize co-evolution and sequence alignment constraints and free-energy minimization to predict probable tertiary folds. The EV-fold model as reported in Lukose *et al.*, 2015, was computed for *C. jejuni* PglC omitting the N-terminal helix. The Rosetta model, while encompassing both soluble and membrane-associated domains, was

calculated for a homolog from the WbaP subfamily, WcaJ. A third model, that of *C. concisus* PglC was calculated by the deep-learning protein-structure prediction algorithm RaptorX (49) by Lingqui Luo in our laboratory. This algorithm does not employ any co-evolutionary or sequence constraints to predict the fold, but rather performs *de novo* structure prediction through trained prediction of residue pair contacts without a template, or homologous, sequence. When superimposed, it is clear the three computational models possess the same overall tertiary structure with some slight differences in the positioning of the loops and helices (Figure 2.9A). This is reflected in the root mean square deviation (RMSD) values for alignment of the RaptorX and Rosetta models with the published EVfold model of 5.6 Å and 5.2 Å, respectively. Superposition of the final model of I57M/Q175M PglC with the three computational models provides validation of the overall tertiary fold observed experimentally (Figure 2.9B). Notably, the greatest divergence between the final experimental model and computationally-derived models of PglC is an ~32 ° rotation opening of the extended loop structure between β-strand 2 and helix D. This divergence could point to a functionally important role for this loop.

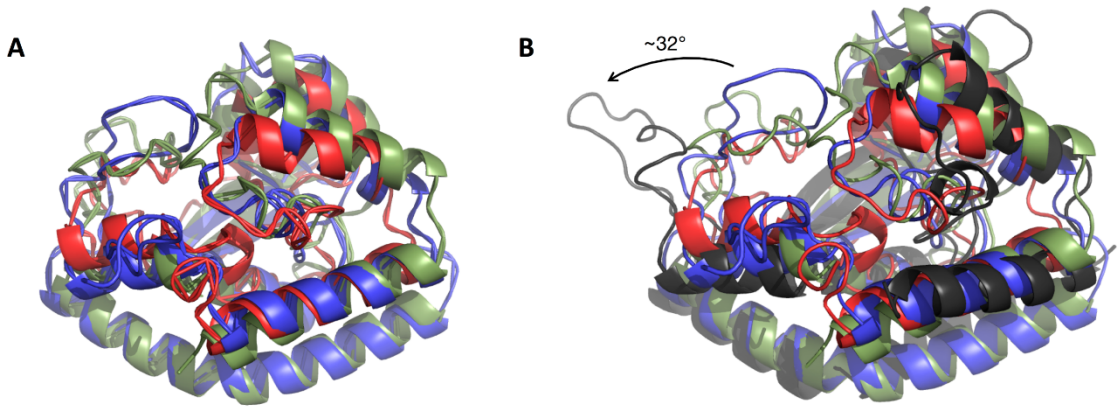

Figure 2.9. Experimentally determined tertiary fold of PglC agrees reasonably well with the computationally derived models. A, Superposition of the three computationally derived models calculated for PglC. EV-fold model of C. jejuni soluble domain shown in red. Model of WcaJ from E. coli computed by Rosetta shown in green. Model of WT *C. concisus* PglC computed by RaptorX shown in blue. **B,** Superposition of the computational models of PglC with the experimentally determined structure (black). Arrow indicates the relative movement of the extended loop joining strand 2 to helix C observed experimentally vs. calculated. RMSDs between the experimental PglC and RaptorX, Rosetta, and EV-fold models were 5.9 Å, 3.1 Å, and 6.1 Å, respectively.

2.4 Conclusions

Determination of the structure of PglC was circuitous and involved the iterative application of various crystallographic methodologies. The final structure required three distinct data sets each providing critical information: the 3.11 Å WT Se-Met dataset allowed initial phasing by SAD, the 2.59 Å I57M/I87M dataset allowed for the molecular replacement and automatic and manual model-building, and the 2.74 Å dataset provided the most complete data for final

structure refinement. Overall, the structure of PglC described in this chapter represents the first example of a novel protein fold and the structure of a monotopic PGT with its membrane-interaction domain. Monotopic proteins are vastly underrepresented in the PDB, representing only 0.05% of known structures. The molecular knowledge gained from the structure of PglC is thus not only applicable to the 15,000 members of the monotopic PGT superfamily, but also to monotopic proteins in general.

CHAPTER THREE

Molecular Structure of PglC Reveals Novel Architecture for Monotopic Membrane Association

The work presented in this chapter is included in the following accepted article: Ray, L.C., Das, D., Entova, S., Lukose, V., Lynch, A.J., Imperiali, B., and Allen, K.N. Membrane association of monotopic phosphoglycosyl transferase underpins function. *Nature Chemical Biology* (2018). *(in press).*

3.1 Introduction

Although membrane proteins make up approximately 30% of the proteome (59), they are vastly underrepresented in the PDB with only 5% of protein structures in the non-redundant PDB representing membrane proteins. These proteins can be categorized by their associations with the membrane both functionally by their association with the membrane and topologically by the physical arrangement of their structure with respect to the membrane (Figure 3.1). Functionally, membrane proteins are classified as either peripheral or integral. Peripheral membrane proteins are water soluble and often associate with lipid bilayers reversibly. Conversely, integral membrane proteins are permanently resident in the membrane, and generally cannot be extracted from the membrane without the addition of detergents (60, 61).

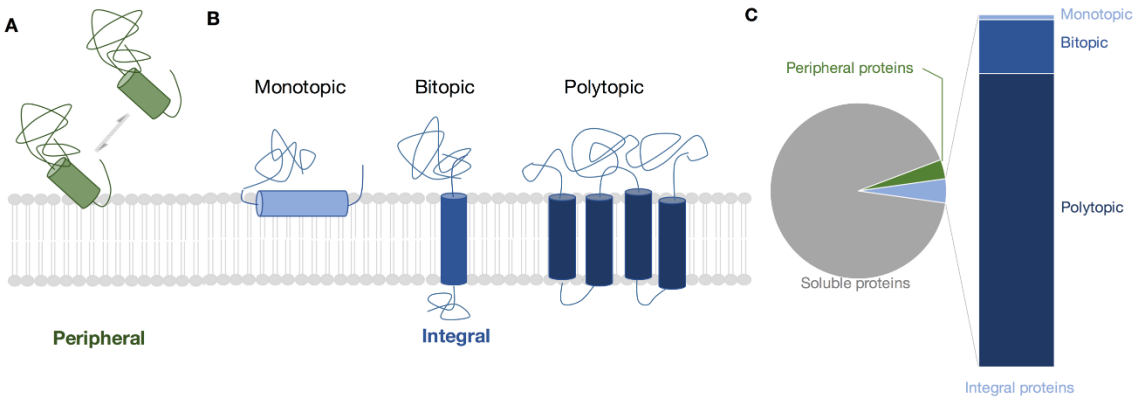

Figure 3.1. Classifications and structural representation of membrane proteins. **A,** Peripheral proteins associate reversibly with the lipid bilayer. **B,** Integral membrane proteins are classified as monotopic, biotopic, or polytopic dependent upon their membrane-associated topology and domain distribution across the membrane. **C,** Distribution of membrane protein structures within the PDB. Numbers for each class compiled from non-redundant (70% sequence ID) PDB, OPM, and MPStruct databases: soluble proteins- 34,987; peripheral proteins – 1,350; integral proteins – 1,679; monotopic – 24; biotopic – 256; polytopic – 1,399.

Integral membrane proteins are further classified as described by Blobel into three topologically distinct categories: monotopic proteins, which associate with only one leaflet of the lipid bilayer, bitopic proteins, which span the entirety of the bilayer once with soluble domains on either side, and polytopic proteins which span the membrane many times with soluble domains on one or both sides of the membrane (62). Not only are the membrane-embedded domains topologically different, the soluble domains and overall structures are largely dissimilar, as well. Although the three-dimensional structures of monotopic and

bitopic proteins are largely unknown, primary structure homology indicates that the soluble domains sample vast areas of known protein fold space. Recent analysis of the metazoan proteome has established that may of the bitopic proteins arose from recruitment of soluble domains to the membrane during evolution (63, 64). Unlike the diversity of folds present for soluble domains, analyses of the integral membrane proteins with known structure reveals membrane embedded domains are largely α-helical in character (65, 66).

3.1.1 Structural and sequence motifs for monotopic-membrane protein interaction

Overall, the nature of the interaction interface for monotopic membrane proteins is markedly less well understood than protein-membrane interactions for peripheral and polytopic proteins; however, a few main themes are prevalent within these interactions. Unsurprisingly, a survey of the role of hydrophobic interactions in membrane association of peripheral and integral monotopic proteins identified amphipathic helices as a highly prevalent structural motif driving membrane association (65). Additionally, bioinformatic analyses suggest basic residues (K, and R) and hydrophobic (L, W, Y, F, L, I, and V) play a key role in anchoring the monotopics (67, 68). Basic residues are thought to satisfy electrostatic interactions with the negatively-charged headgroups of the

membrane phospholipids. As expected, based on molecular dynamics (MD) simulations and structural bioinformatics, hydrophobic interactions anchor these monotopic proteins to the hydrophobic core of the membrane. Overall, it is apparent that there are two subclasses of monotopic proteins, those with shallow insertions that interact primarily with the bilayer surface, and those that penetrate deeper into the hydrophobic core of the membrane. MD simulations have noted that the membrane association of the monotopics which extend deeper into the membrane result in local perturbations in the lipid bilayer (67), which could be of functional importance to allow access to hydrophobic, membrane-embedded substrates.

3.1.2 Monotopic membrane protein membrane-interaction modes

It is universally acknowledged that integral monotopic membrane proteins associate with the membrane permanently, with their soluble domains not traversing the bilayer as seen in bitopic and polytopic proteins; however, there is considerable ambiguity and dispute within the literature as to what constitutes integral monotopic proteins. Despite the ambiguity, five main classes of membrane-interaction modes are evident (61, 66, 67, 69). These modalities of monotopic membrane protein interaction with membrane comprise amphipathic

helices parallel to the membrane plane, single-pass TM containing proteins with a soluble domain on only one face of the membrane, hydrophobic loops extending into one leaflet, hydrophobic patches coupled with electrostatic interactions, and covalently-bound lipid anchors. The membrane-interaction modality of the monotopic proteins of known structure can all be classified into one of these five modes (Figure 3.2).

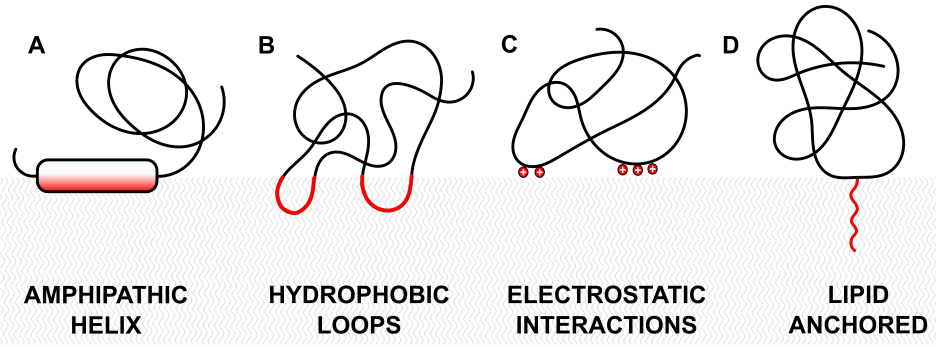

Figure 3.2. Schematic representation of interactions between monotopic membrane proteins and the membrane. A, Interaction via an amphipathic α-helix parallel to the plane of the membrane. **B,** Interaction via hydrophobic loops. **C,** Electrostatic or ionic interactions with lipid head groups. **D,** membrane association via a covalent fatty acyl, prenyl or glycosylphosphatidylinositol post-translational modification.

PglC was predicted to adopt the canonical "lollipop"-like topology of a single-pass TM helix containing monotopic protein from secondary structure and hidden Markov model TM helix prediction (20); however, the molecular structure presented in chapter 2 does not appear to adopt this membrane spanning

topology. Instead, the N-terminal hydrophobic helix is kinked resulting in both N- and C- termini residing within the same plane. Comparison of the structure of PglC with the canonical membrane interaction motifs reveals that PglC does not appear to fit. Thus structural, biochemical, and bioinformatics approaches were used to understand the mode of membrane-interaction in PglC. The idiosyncratic kinked reentrant membrane helix of PglC represents a novel architecture for membrane interaction. However, due to the packing of PglC within the crystal lattice mediated through interdigitation of the kink-forming proline residues, further evidence is required to convincingly assign this unique structure as functionally and physiologically relevant.

3.2 Materials and Methods

3.2.1 Covariance mapping

The webserver GREMLIN (http://gremlin.bakerlab.org/) was used to create a statistical model of the conservation and covariance in the PglC family alignment by Lingqi Luo (Allen Lab, BU) as previously reported (17). Contact pairs with greater than or equal to 99% probability of co-evolution were plotted as pseudo-bonds with UCSF Chimera onto the PglC structure for analysis.

3.2.2 Hydrophobicity analysis

Hydrophobicity of residues of PglC was analyzed in PyMol according to the Eisenberg normalized consensus hydrophobicity scale (70). Amphipathic α-helices identified from hydrophobic coloring in PyMol were analyzed via helical wheel projections created using the Helixator (71) webserver (http://www.tcdb.org/progs/helical_wheel.php). The following helix sequences were used in construction of the projections: helix C – KFGKLMRSL (Residues 86-94), helix D – LDELPQLFNVLK (Residues 96-107), helix I – FMLDVKIALQTIEKVLK (Residues 170-186).

3.2.3 Calculation of $\Delta G_{transfer}$ and PglC placement within the membrane

The free energy of transfer of PglC into the membrane was calculated as $\Delta G_{transfer}$ = -42.9 kcal/mol with the PPM (Positioning of Proteins in Membrane) server (http://opm.phar.umich.edu/server.php) (72). The core of this calculation is a summation of two terms – a solvent accessible surface-area dependent term and an electrostatic term. The solvent accessible surface-area term accounts for van der Waals interactions, hydrogen-bonding, solvent-solute interactions, and the entropy of the solvent in the hydration shell around the protein. In addition, the electrostatic term includes consideration of solvation energy of dipoles and ions

coupled to penalties for placement of ionizable groups in hydrophobic environments.

3.2.4 Electrostatic surface analysis

Electrostatic surface analyses for PglC were performed using the Adaptive Poisson-Boltzmann Solver (APBS) (73) plug-in for PyMol (Fig. 3c and 3d). APBS was run in Linearized Poisson-Boltzmann Equation mode with surface calculation by cubic B-splines with harmonic average smoothing. Electrostatic surface was visualized with contours at ± 5 kT/e.

3.2.5 Helix D geometry analysis

The change of the N-terminal end of helix D from alpha-helical to 3_{10}- helical geometry was confirmed using the webserver HELANAL-plus (74). Helix D geometry was analyzed in the following partitions: full-length (residues 92-104), N-terminus (residues 92-96), C-terminus (aa 96-104).

3.2.6 Structure packing analyses

UCSF Chimera (75) was used to calculate the solvent-accessible surface area (SAS) and the volume contained within, or solvent-excluded volume (SEV), for full-length PglC, as well as for the membrane embedded (intra-membranous) and

extra-membrane portions of PglC and using the membrane plane defined by hydropathy analyses. Solvent accessible surfaces used in SAS and SEV volume calculations. The surface-to-volume ratio (SVR) was defined by the following equation: SVR = SAS/SEV. Surface-to-volume ratios (SVRs) calculated for full-length PglC, intra-membranous, and extra-membranous domains of PglC. A test-set of ten proteins of similar molecular weight to PglC and determined at resolutions between 4 Å - 1.5 Å was chosen from the PDB to compare PglC to other proteins of known structure. Test-set PDB IDs: 5LDO, 5KBC, 4L1N, 4HVR, 3USY, 3CGN, 3ZNB, 2IQI, 1GZE, 3DO9.

3.3 Results and Discussion

3.3.1 PglC N-terminal hydrophobic helix adopts distinct re-entrant topology

The molecular structure of PglC, as presented in chapter 2, is the first experimentally determined structure of monotopic PGT. PglC adopts a novel protein fold where the N-terminal hydrophobic helix, predicted to transverse the membrane, is instead kinked into a helix-break-helix motif. This helix mediates the formation of a tetrameric crystal-lattice contact through interdigitation of proline residues at the kink from neighboring symmetry mates (Figure 2.8B). As the molecular structure of PglC did not assume the “lollipop”-like topology expected

for a predicted single-pass transmembrane protein, detailed structural analyses were undertaken to gain insight into the potential membrane interaction interface of PglC. As such, the analyses aimed at determining the functional relevance of the observed helix-break-helix structure of the N-terminal hydrophobic helix.

The amphiphilic nature of the phospholipids which compose lipid bilayers in biological systems imparts a defined topology to the membrane as a whole. The phospholipids are arranged such that the hydrocarbon tails form the hydrophobic core, and hydrophilic head-groups form the membrane surface. As a result, membrane-proteins often have asymmetrical distribution of hydrophilic and hydrophobic residues to create unique intra-membranous and extra-membranous surfaces. Mapping the normalized consensus hydrophobicity(70) of each residue of PglC illustrates a clear plane of hydrophobicity bounded by helices C, D, and I that includes the kinked, hydrophobic helix (A, B) (Figure 3.3A). While it was evident that the N-terminal helix would display a high frequency of hydrophobic residues based on primary structure, the amphipathic nature of helices C, D, and I became apparent upon mapping of hydrophobicity to the structure. Helical wheel projections of helices C, D, and I confirm clear amphipathic character with one face of the helix surface marked by hydrophobic residues opposite a more hydrophilic surface marked by residues with polar groups (Figure 3.3B). The

residue hydropathy analyses suggest the membrane-interaction interface for PglC would bury the hydrophobic helix-break-helix within the membrane forming a reentrant membrane helix (RMH) resulting in both N and C-termini of the helix resident on the same side of the membrane. The resulting membrane-interaction interface for PglC would be bounded by the three co-planar amphipathic helices embedded in the surface of the membrane.

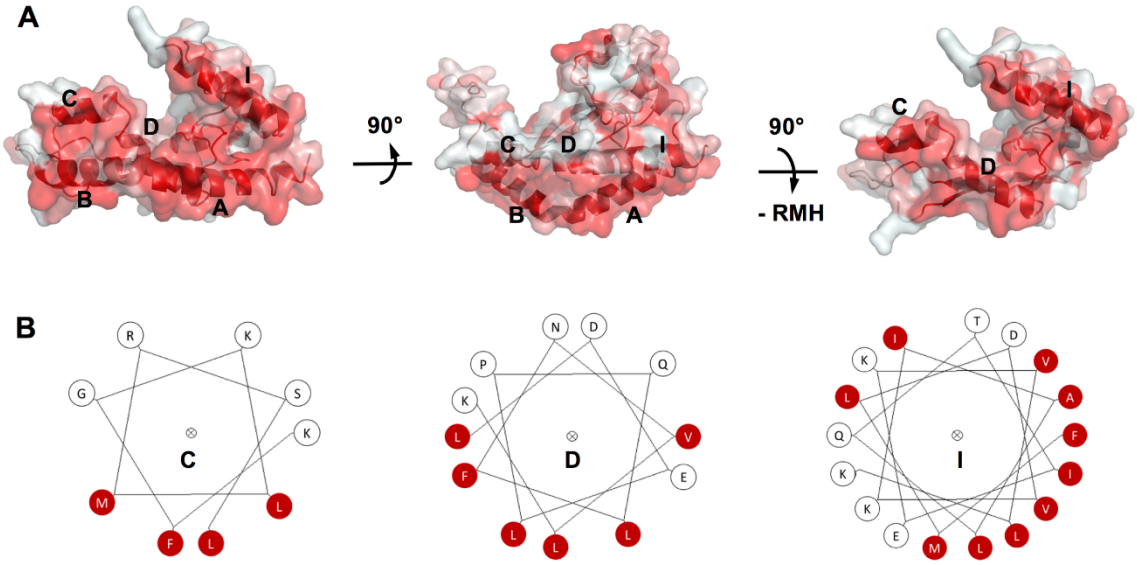

Figure 3.3. Hydrophobicity and electrostatics corroborate membrane placement. **A,** The hydrophobic surfaces (ramped from white to red for polar to hydrophobic) of amphipathic helices labeled C, D, and I establish a planar hydrophobic surface (left); also shown rotated by 90° with RMH removed (right). **B,** Helical wheel representations for helices C, D, and I illustrate the clear amphipathic nature of the co-planar helices. Hydrophilic residues represented in white, hydrophobic residues represented in red.

Independent confirmation of the positioning of PglC within the membrane was accomplished through the use of the Positioning of Proteins in Membrane (PPM) server (72). Assignment of the membrane-embedded position via this server is irrespective of the qualitative assignment of the hydrophobic plane from residue hydrophobicity mapping. The PPM server calculates the optimal rotational and translational orientation of proteins within membrane through minimization of the free energy, $\Delta G_{transfer}$, of the transfer of the protein from water to a simulated, homogeneous DOPC (1,2-Dioleoyl-sn-glycero-3-phosphocholine) membrane slab. Minimization of $\Delta G_{transfer}$ for the coordinate set of PglC resulted in the placement of PglC in an orientation in which the kinked RMH extends 14 Å deep into the membrane and the three amphipathic helices (AH) identified through hydrophobicity analysis (C, D, and I) sit at the membrane interface (Figure 3.4). It should be noted that the discrepancy between positioning of the membrane planes calculated by the PPM server and estimated by hydrophobicity analyses of the RMH and AHs suggests that PglC may be positioned approximately 5 Å deeper in the membrane than calculated by the PPM server (Figure 3.5). The inconsistency arises from the modelling of the bilayer by the PPM server, where the membrane is treated as a rigid hydrophobic slab which neglects protein-head group interactions, as well as the intrinsic deformability of

the membrane (67). The membrane-bound orientation of PglC calculated by the PPM server represents the minima of $\Delta G_{transfer}$ = -42.9 kcal mol^{-1}, which places 27% of the residues of PglC embedded within the membrane.

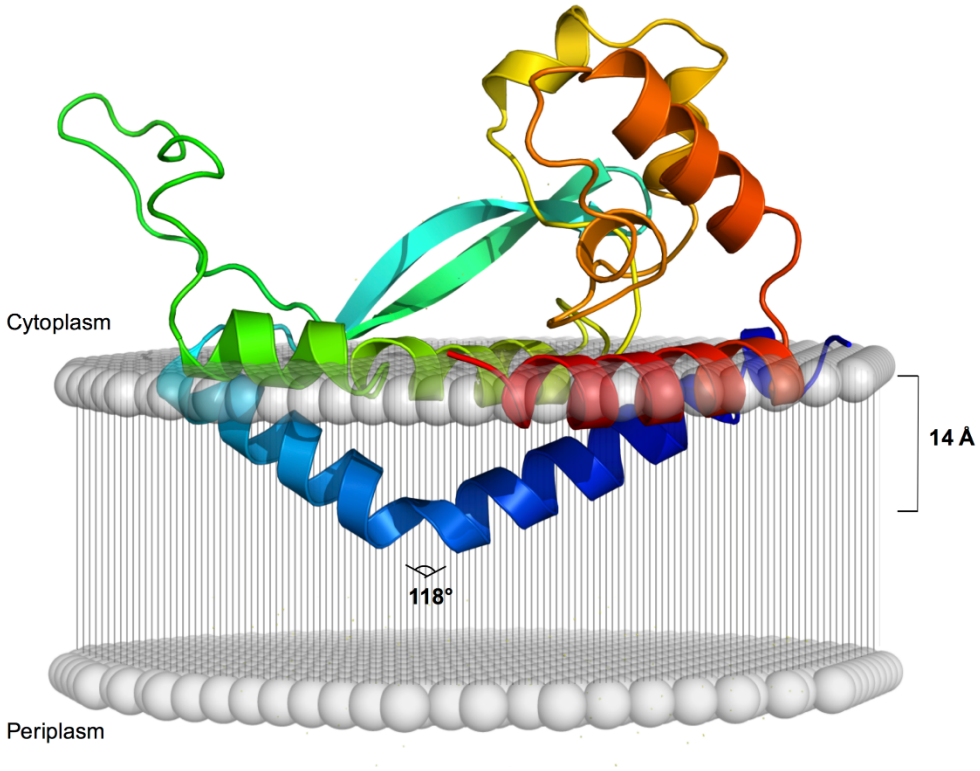

Figure 3.4. PglC reveals a distinct architecture and topology for monotopic membrane proteins. Predicted position of PglC with respect to the membrane including the reentrant membrane helix (RMH) (N- to C-terminus colored blue to red).

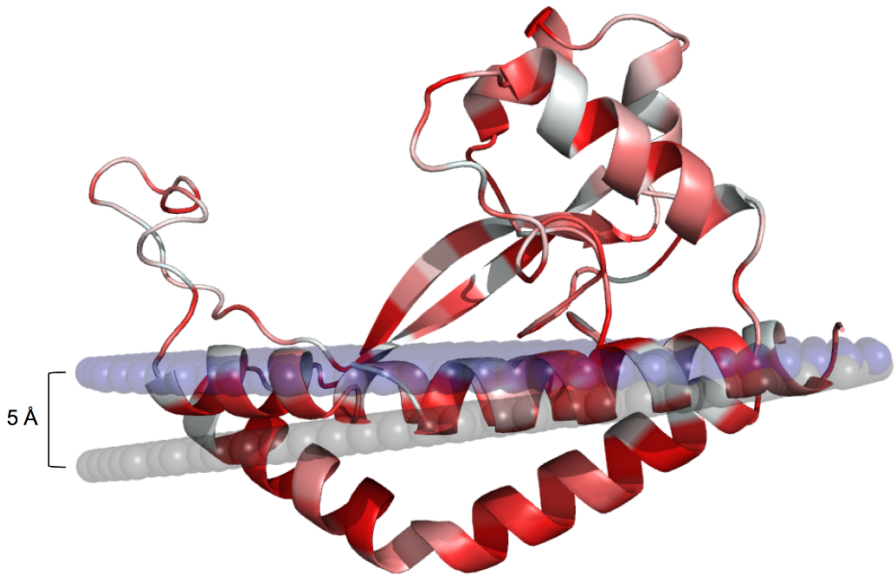

Figure 3.5. Comparison of PglC membrane plane location calculated by PPM server (gray spheres) and manually-positioned plane (blue spheres) using hydropathy analyses of the RMH and coplanar α-helices. Model of PglC rendered with hydrophobic coloring.

The extensive membrane interaction interface observed for PglC is unusual compared to other previously-reported monotopic membrane protein structures. In contrast, the known structures of monotopic membrane proteins are principally comprised of extensive soluble domains and comparatively minimal membrane-associated domains as exemplified by the well characterized, classical examples of monotopic proteins: prostaglandin H2 synthase (PGHS, PDB: 1CQE) (76), fatty acid acyl hydrolase (FAAH, PDB:1MT5) (77), and squalene synthase (PDB: 2SQC) (78) (Figure 3.6).

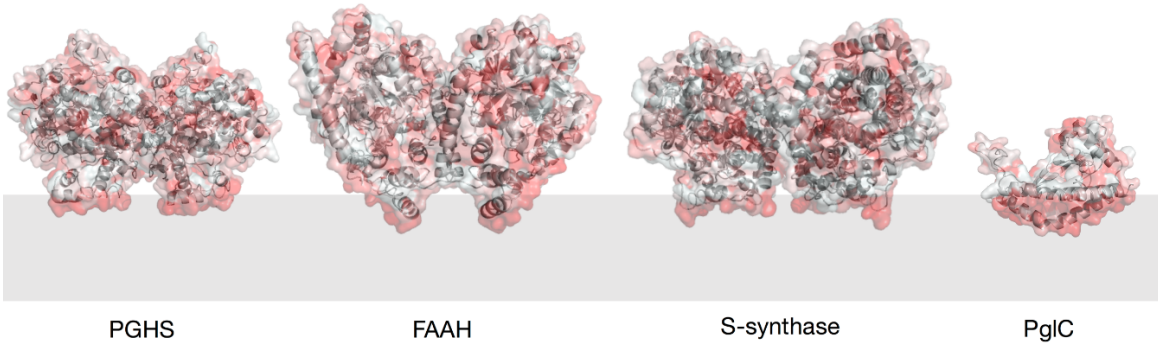

Figure 3.6. Membrane interaction surfaces of exemplar monotopic proteins. Protein surface colored by normalized consensus hydrophobicity from hydrophobic (red) to hydrophilic (white). Membrane represented by gray box. Location of membrane plane calculated with PPM server (http://opm.phar.umich.edu/server.php) for 1CQE (PGHS), 1MT5 (FAAH), 2SQC (S-synthase), and PglC.

The extensive soluble domains of these monotopic examples comprise known soluble folds (e.g. Rossmann-like, α-β-α sandwich-like) which have been recruited to the membrane interface. In addition, computation of the membrane-embedded interface and the free energy of transfer into the membrane for all monotopic proteins of known structure provides a quantification of the divergence in membrane association mode for known monotopic structures and PglC. On average, monotopic proteins currently in the non-redundant PDB have comparatively minimal interaction interfaces with only 3.9% of their residues embedded and a significantly less energetically favorable interaction with the

membrane illustrated by an average $\Delta G_{transfer}$ = -17.4 kcal mol^{-1} (Table 3.1). It is worth noting that there are some examples of monotopics with interaction interfaces which appear more like that of PglC quantitatively (Alternative oxidase- PDB: 3VV9, 16% embedded, $\Delta G_{transfer}$ = -61.4 kcal mol^{-1}; Monofunctional glycosyltransferase- PDB:3VMT, 11% embedded, $\Delta G_{transfer}$ = -35.4 kcal mol^{-1}, Figure 3.7); however, on average, the membrane interaction interfaces are considerably smaller.

Table 3.1. Quantitative comparision of membrane interaction interfaces for all monotopics of known structure, including PglC (red).

Macromolecule	PDB ID	ΔGtransfer (kcal mol)	Total AA in chain	Embedded AA	% embedded	% exposed
Cyanide-insensitive alternative oxidase	3VV9	-61.4	329	54	16.4	83.6
Prostaglandin H2 Synthase	1CQE	-43.4	580	44	7.6	92.4
PglC	**5W7L**	**-39.5**	**200**	**47**	**23.5**	**76.5**
Cyclooxygenase-2	1CX2	-37.4	587	56	9.5	90.5
Lanosterol 14-alpha demethylase	4LXJ	-36.1	536	43	8.0	92.0
Monofunctional glycosyltransferase	3VMT	-35.4	263	29	11.0	89.0
Fatty acid alpha-dehydrogenase	4HHS	-33.7	652	38	5.8	94.2
Lysophosphatidic acid acyltransferase	5KYM	-32.4	247	40	16.2	83.8
Squalene-hopene cyclase	2SQC	-27.8	631	10	1.6	98.4
Monoamine oxidase A (human)	1O5W	-23.1	534	37	6.9	93.1
Penicillin-binding protein 1B	3VMA	-21.7	758	20	2.6	97.4
Monoamine oxidase A (rat)	2BXR	-21.4	527	35	6.6	93.4
WaaA	2XCI	-20.3	374	2	0.5	99.5
Sulfide:quinone oxidoreductase	3HYW	-19.4	430	24	5.6	94.4
Flavoprotein-ubiquinone oxidoreductase	2GMH	-18.5	584	7	1.2	98.8
15-lipoxygenase	1LOX	-14.3	662	6	0.9	99.1
LpxK	4EHW	-13.9	317	7	2.2	97.8
Retinoid isomerohydrolase P65	4F3D	-13.9	533	2	0.4	99.6
Type II NADH dehydrogenase	4NWZ	-13.7	405	2	0.5	99.5
NAPE-PLD	4QN9	-11.3	549	18	3.3	96.7
Peptidoglycan glycosylatransferase	2OQO	-11	200	12	6.0	94.0
Wall teichoic acid polymerase (TagF)	3L7I	-10.8	729	18	2.5	97.5
NADH-ubiquinone	4G6G	-10.6	502	11	2.2	97.8
15-lipoxygenase-2	4NRE	-10.6	696	9	1.3	98.7
Retinoid isomerohydrolase P65	3FSN	-10.3	533	10	1.9	98.1

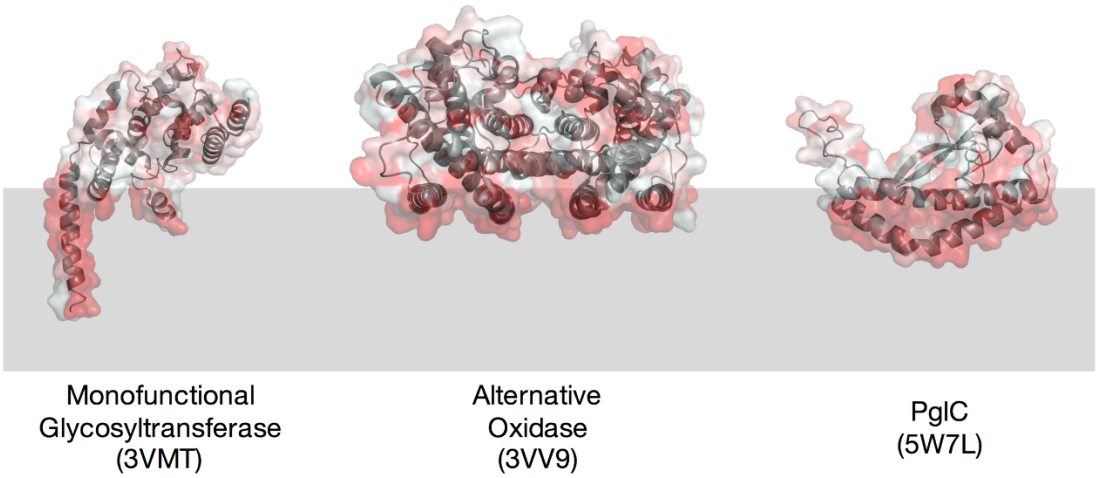

Figure 3.7. Membrane interaction surfaces of other highly-embedded monotopic proteins. Protein surface colored by normalized consensus hydrophobicity from hydrophobic (red) to hydrophilic (white). Membrane represented by gray box. Location of membrane plane calculated with PPM server (http://opm.phar.umich.edu/server.php) for 3VMT (16% embedded) , 3VV9 (11% embedded), and PglC (27% embedded).

3.3.2 Biochemical and bioinformatic validation of PglC membrane-interaction mode

The reentrant-helix topology of the hydrophobic helix at the core of the membrane-interaction mode in PglC appears unique across other monotopic membrane proteins of known structure. This suggests the RMH motif in PglC is the first-in-class example of a novel modality of membrane interaction. Bioinformatic and biochemical analyses of PglC corroborate the observed membrane interaction modality as the physiologically-relevant conformation for PglC. Notably, the crystallographically observed topology is consistent with biochemical studies on a related monotopic PGT superfamily member, WcaJ,

which propose this re-entrant topology for the hydrophobic domain homologous to PglC (34% sequence identity) (79).

Bioinformatic analysis supports the reentrant topology through calculation of covariance and co-evolution of residue pairs within the monotopic PGT-superfamily sequence alignment. Covariance and co-evolution analyses infer that two residues in physical contact will co-vary throughout evolutionary space, and therefore throughout the primary sequences, to retain the physical interaction (80). Thus, a high probability of covariance between pairs of residues correlates with a high-probability of physical contact. In PglC, ten pairs of residues with 99% probability of covariance map to contacts between the RMH and the planar amphipathic helices (Figure 3.8). These pairs include π-π stacking, van der Waals, and hydrogen-bonding interaction modes.

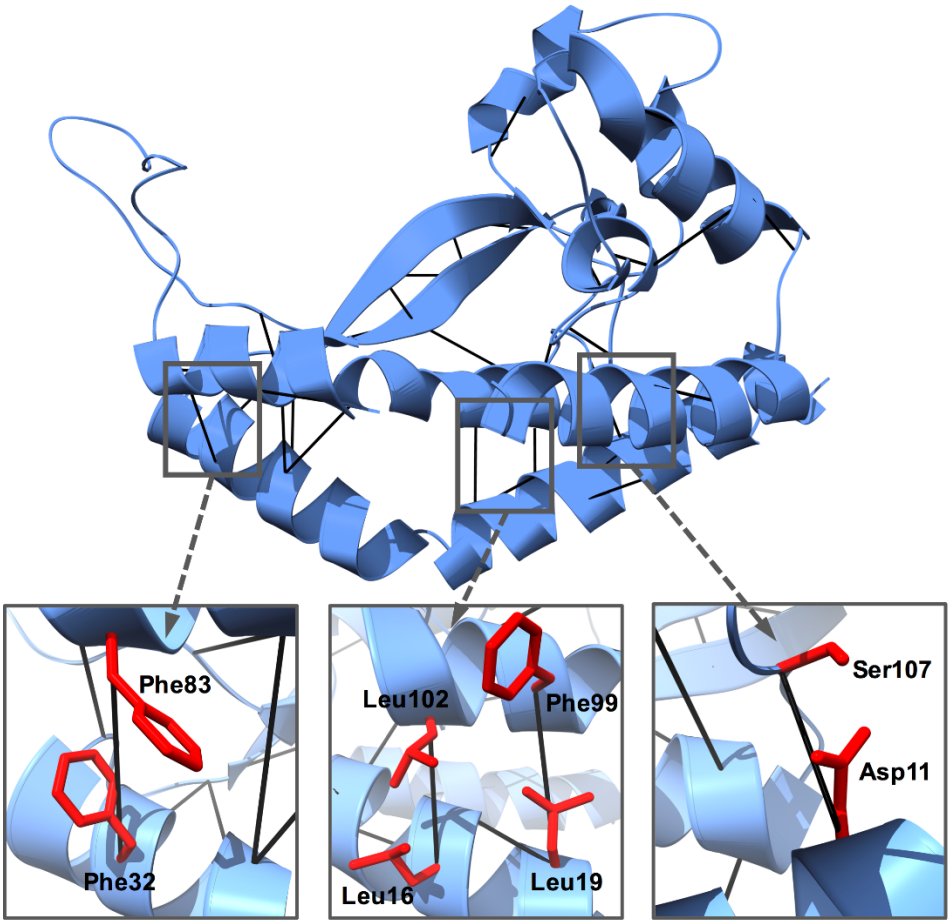

Figure 3.8. PglC crystallizes in a native conformation. Co-evolution and covariance analyses across the monotopic PGT superfamily predict 10 interactions with ≥99% probability of contact (black lines). For example, Phe32/Phe83 exhibit π-π-stacking, Leu19/Phe99; Leu16/Leu102; Thr29/Leu92; Tyr34/Val41; Ala30/Val41 make hydrophobic interactions between the RMH and the coplanar helices and Asp11/Ser107 form a hydrogen bond.

In concert, the reentrant topology was validated for PglC via substituted cysteine accessibility method (SCAM) (81, 82) analyses performed by Sonya Entova (Imperiali Lab, MIT). These experiments illustrate that cysteines

engineered at the N- and C-termini of PglC both display cytoplasmic localization (Figure 3.9). As a result, the kinked helix must also adopt reentrant rather than membrane spanning topology in whole cells. Additionally, highly-conserved basic residues at the N-terminus of PglC, and therefore the reentrant helix, are consistent with the "positive-inside" rule for a topology that faces the cytoplasm (83).

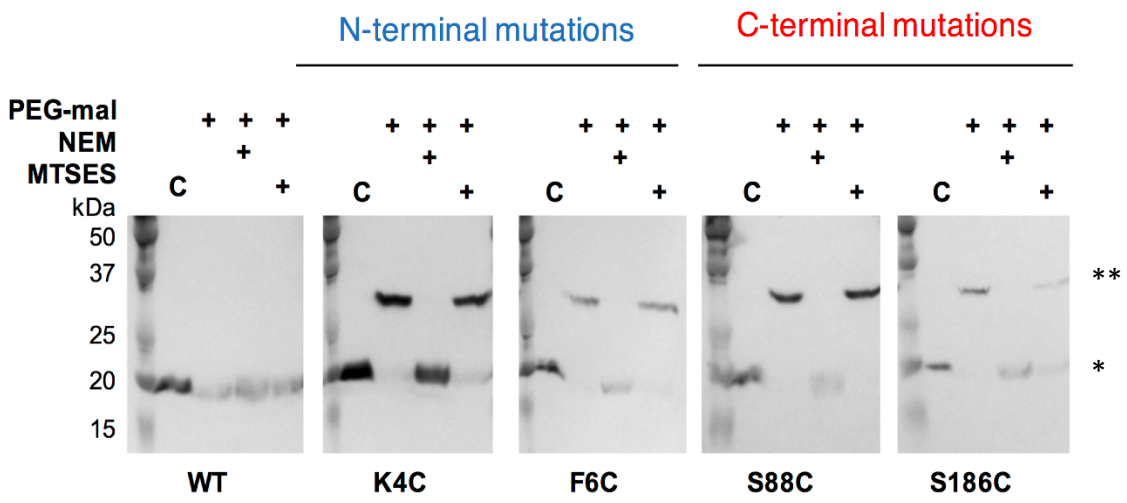

Figure 3.9. PglC crystallizes in a native conformation. *In vivo* SCAM analysis indicates that the N-and C-termini of the RMH are localized at the cytoplasmic face (* = native PglC; ** = PglC labeled with PEG-mal, C = control, no PEG-mal labeling). PEG-maleimide (PEG-mal); N-Ethylmaleimide (NEM); 2-sulfonatoethyl methanethiosulfonate (MTSES). Data and figure courtesy of Sonya Entova (Imperiali Lab, MIT).

In addition to the bioinformatics and SCAM confirmations of PglC membrane topology, observation of an endogenous PE lipid head group in the electron density localized at the predicted membrane interaction interface

provides further evidence for placement of the membrane plane (Figure 3.10A). As PE was not added during at any point during purification or crystallization, the presence of PE within the crystal could only have resulted from retention of tight-binding, endogenous PE throughout purification and crystallization of PglC. To confirm the presence of PE in protein samples used for crystallization, thin layer chromatography (TLC) analyses were performed by Dr. Debasis Das (Imperiali Lab, MIT). TLC demonstrates the presence of endogenous POPE (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine) and POPG (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoglycerol) in approximately a > 3:1 stoichiometry (Figure 3.10B). The retention of endogenous lipids through purification is common across membrane protein structures with resolved lipids (84). The composition of the binding-site with a hydrogen bond between a positively-charged, arginine side chain (Arg8) and the PE phosphoryl moiety, is consistent with the types of interactions observed for PE-protein binding sites across other membrane proteins (84) (Figure 3.10A). The localization and binding-mode of the PE moiety supports the positioning of the membrane with respect to PglC in our model (Figure 3.10C). It should be noted that the observed PE head-group is not in a physiological orientation with the ammonium group facing towards the predicted hydrophobic core of the membrane, therefore leaving the hydrocarbon tail solvent exposed. The

non-physiological orientation is enforced through a π-cation interaction (85) between the -NH_3^+ and the aromatic ring of a phenylalanine (Phe35) residue from a symmetry-related molecule in the crystal lattice (Figure 3.10C). Although TLC analysis suggests a 3:1 mixture of PE:PG present in the crystallization sample, the shape and fit of the electron density and clear cation-π interaction with the symmetry mate Phe35 corroborate assignment of the density as a PE molecule rather than PG (Figure 3.10D).

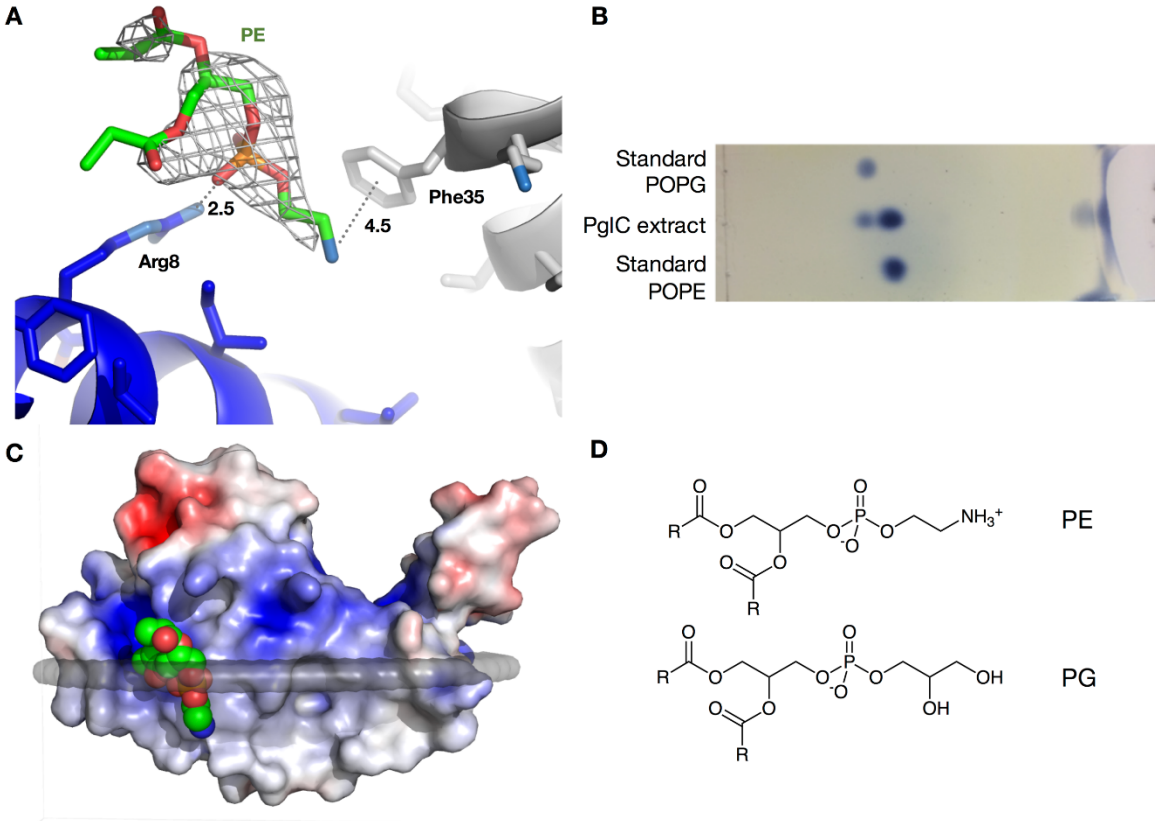

Figure 3.10. Crystallographically-observed PE head group corroborates position of the membrane-interaction interface. A, F_o-F_c simulated annealing omit map contoured to 3σ shown as gray wire mesh. PE molecule 303 (chain A) shown in green sticks. RMH of PglC chain A shown in blue. Phe35 from a symmetry mate shown in gray. Interactions between the PE phosphoryl moiety and Arg8 (hydrogen bond) and Phe35 (cation-π) shown with distances given in Å. **B,** TLC analysis performed by Dr. Debasis Das (Imperiali Lab, MIT) demonstrates the presence of both phosphatidylethanolamine (POPE) and phosphatidylglycerol (POPG) phospholipids in the purified PglC at a ~3:1 stoichiometry. **C,** PE head group binding site at Arg8 locates this position to the membrane interface. PglC colored by electrostatic potential from +5 kT/e (blue) to -5 kT/e (red). **D,** Chemical structures of the PE and PG head-groups. R groups represent the acyl chains of the phospholipids.

3.3.3 Small tertiary fold motifs underpin the structure and function of PglC

Overall, PglC is largely divergent from other monotopic proteins of known structure in membrane-association interface, free energy of transfer to the membrane, and in general domain structure. The novel architecture of PglC encompassing the unique RMH containing membrane-embedded domain and minimal cytosolic domain, appears largely open with only the double-twisted loop motif forming a typical globular-like domain. Though the structure of PglC appears poorly packed, analysis of surface-to-volume ratio (86, 87) for full-length monomeric PglC reveals an (SVR) of 0.44 $Å^{-1}$ which is similar to that of other globular proteins of comparable size in a test-set of structures from the PDB, where the average SVR = 0.47 $Å^{-1}$ (Figure 3.11). Therefore, the novel architecture as a whole can be considered globular. In addition, computational analysis with Voidoo (88) to determine the presence of solvent-excluded cavities within PglC revealed only small cavities with volumes of 0.554 ± 0.232 $Å^3$ and 0.754 ± 0.344 $Å^3$, respectively (Figure 3.12). The insignificant cavities, with volumes considerably smaller than a water molecule, coupled with the SVR analysis demonstrate that the fold of full-length monomeric PglC is a well-packed globular-like fold despite its open appearance. It follows that PglC must be structurally efficient with the

establishment of its fold in order to perform the PGT reaction with a minimal scaffold.

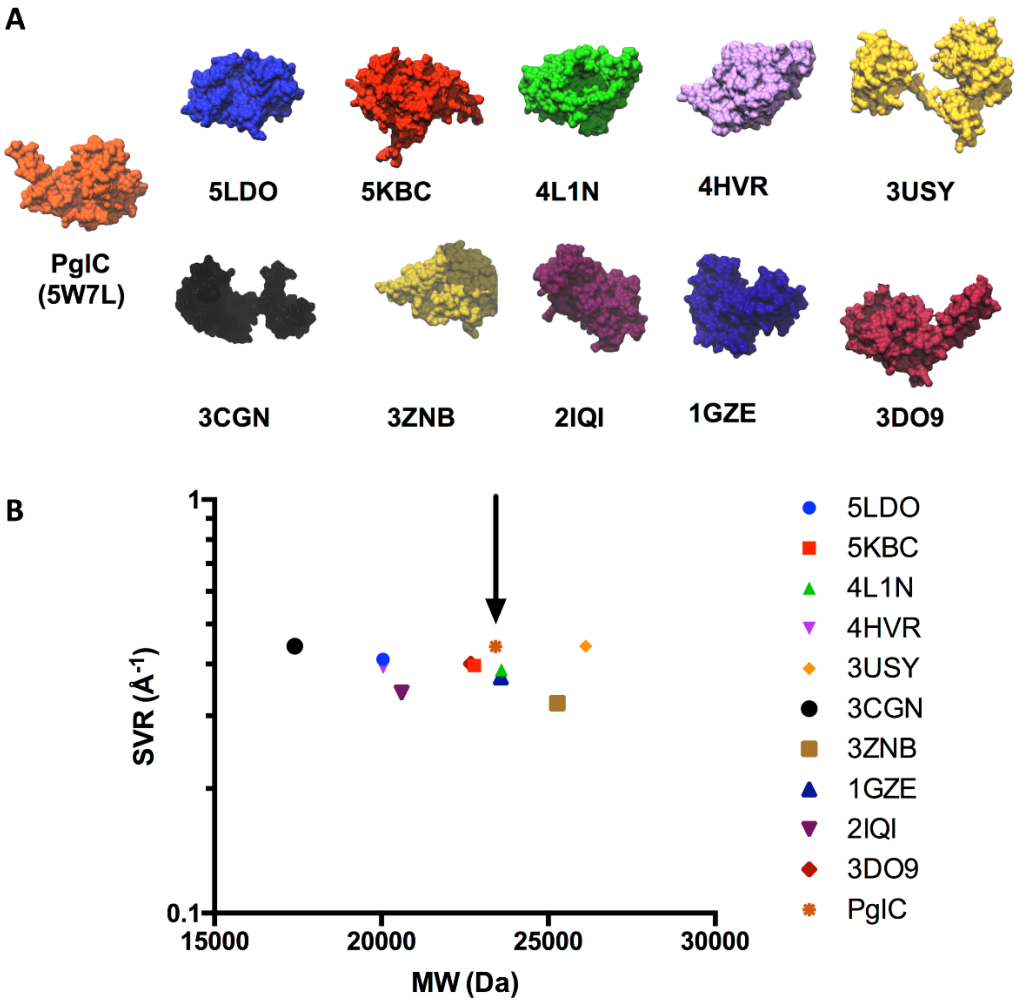

Figure 3.11. Molecular packing of PglC compared to structures of similar size. **A,** UCSF Chimera (75) was used to calculate the solvent accessible surface area (SAS) and the volume contained within, or solvent excluded volume (SEV), for full-length PglC and a test-set of similar proteins from the PDB. The surface-to-volume ratio (SVR) was defined as follows, SVR = SAS/SEV Test set of similar structures from the PDB shown in surface representation. **B,** SVR of PglC (arrow) is similar to those of a test set of proteins of comparable molecular weight from the PDB.

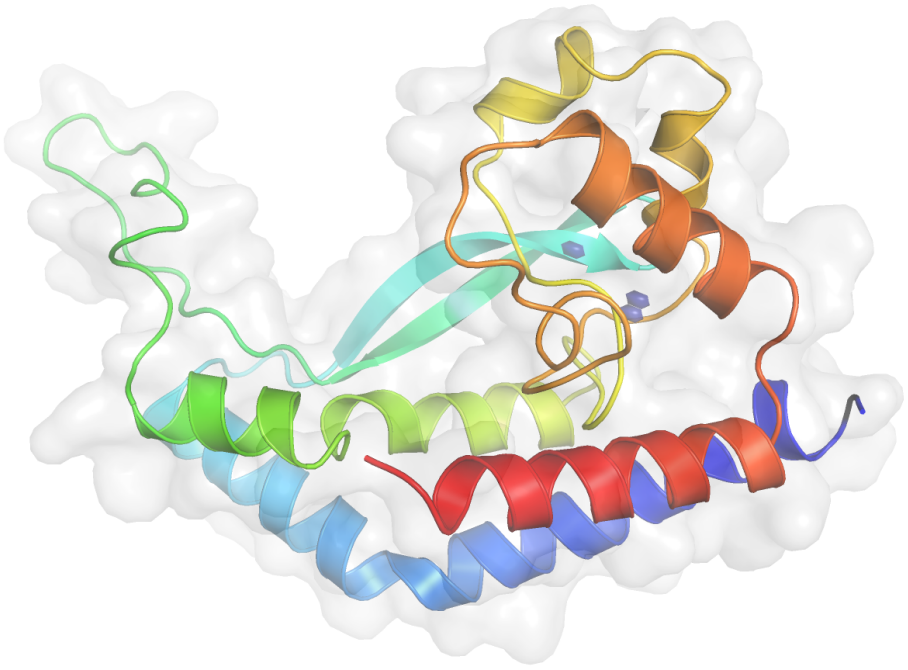

Figure 3.12. Solvent-excluded cavities detected in the double-twisted loop domain of PglC. Cavities detected by Voidoo shown in dark blue with volumes of 0.554 ± 0.232 Å^3 and 0.754 ± 0.344 Å^3, respectively. Solvent-accessible surface of PglC displayed as rendered by PyMol in transparent gray.

Overall, the structure of PglC is parsimonious in that its tertiary structure features are established not through large domain-domain interactions, but rather through short-range non-covalent fold motifs in the form of proline-kinks and elaborate hydrogen-bonding networks. A proline-kink established by Pro24 in tandem with a short intra-molecular hydrogen bond between the Ser23 hydroxyl and the backbone carbonyl of Ile20 establishes the helix-break-helix motif

necessary for the reentrant topology of the RMH (Figure 3.13B). Additionally, β-sheet-like network of short intra-molecular hydrogen bonds between the backbone amide and carbonyl groups of proximal segments stabilize the base of the globular double-twisted loop motif (Figure 3.13C). The RMH N-terminus is pinned into the base of the double-twisted loop-motif through short hydrogen-bonding interactions between a triad of polar residues (Tyr2, Asp169, and, Tyr160). This critical structural feature supporting the reentrant topology is consistent with proposed structural role of absolutely conserved Asp169 determined by mutagenesis studies by Lukose, *et al.* (17) (Figure 3.13D). Furthermore, the strictly conserved Pro-Arg-Pro (111-113) motif in the double twisted loop creates a rigid turn which pins Arg112 towards the active-site catalytic dyad (figure 3.13E). This forced orientation allows the guanidinium side chain to be well-positioned for interaction with the uracil moiety of the UDP-sugar (17).

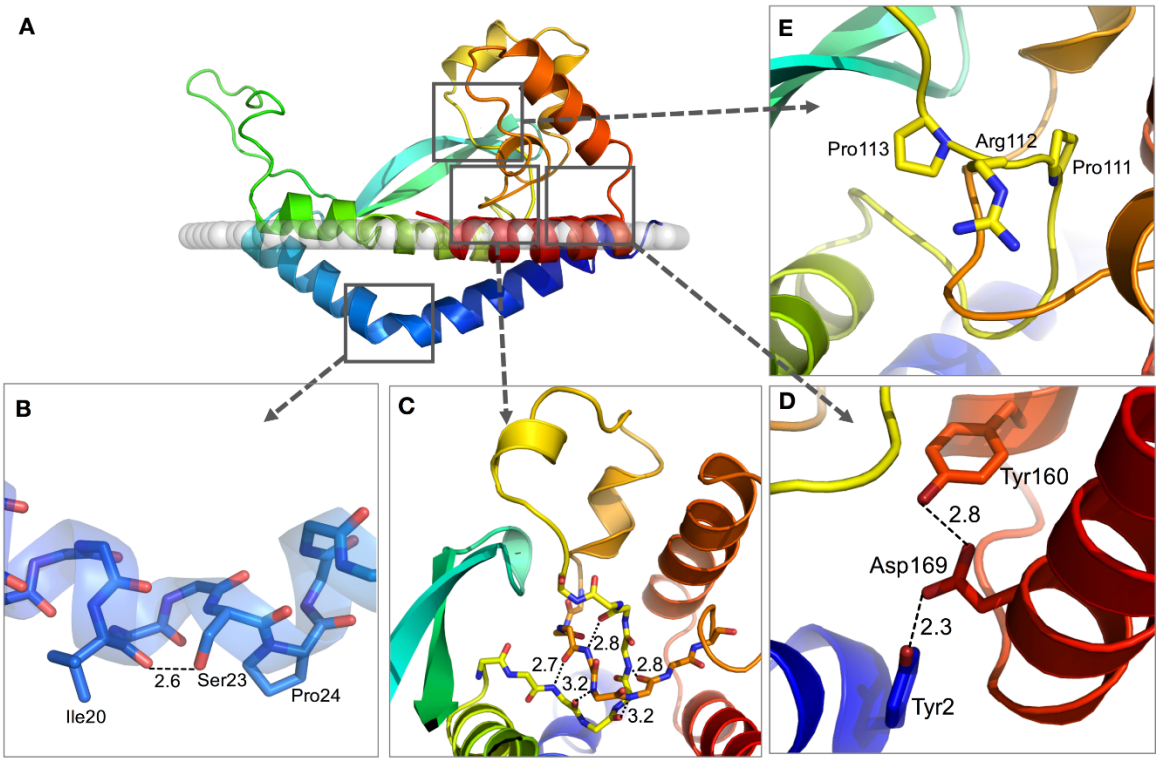

Figure 3.13. Proline-kinks and short hydrogen bonds establish critical interactions that impart unique structure-function relationships in PglC (A). B, A proline-kink from Pro24 in tandem with a short intra-molecular hydrogen bond between Ser23 and the backbone carbonyl of Ile20 establishes the helix-break-helix motif of the RMH. **C,** An extensive hydrogen-bonding network of short intra-molecular hydrogen bonds between backbone amide and carbonyl groups stabilize the double-twisted loop motif. **D,** A triad of polar residues (Tyr2/Asp169/Tyr160) form a hydrogen-bonding network that establishes critical intra-molecular interactions between helices A (blue), H (orange) and I (red). **E,** A strictly-conserved Pro-Arg-Pro (residues 111-113) motif orients Arg112 towards the active site. All hydrogen bond distances given in Å.

3.3.4 The AHABh-motif is structurally and functionally significant

The AHABh-motif (strands 1-2, helices C-D) encompasses both of the structurally efficient fold-motifs seen in PglC: a hydrogen bonding network, and structurally and functionally critical proline kinks. The tertiary fold of the AHABh-motif is established through a network of hydrogen bonds stabilizing the close interaction between β-strand 2 with the full length of helix D (Figure 3.14). The C-terminal end of helix D is held in close proximity to the β-hairpin through interactions between both the backbone amide and tyrosyl hydroxyl of Tyr58 to the side chain carbonyl of Asn100 and Lys103 ammonium group, respectively. Similarly, N-terminus of helix D is stabilized within the AHABh-motif through a hydrogen bond network formed between Asp93, Phe60, and Thr62.

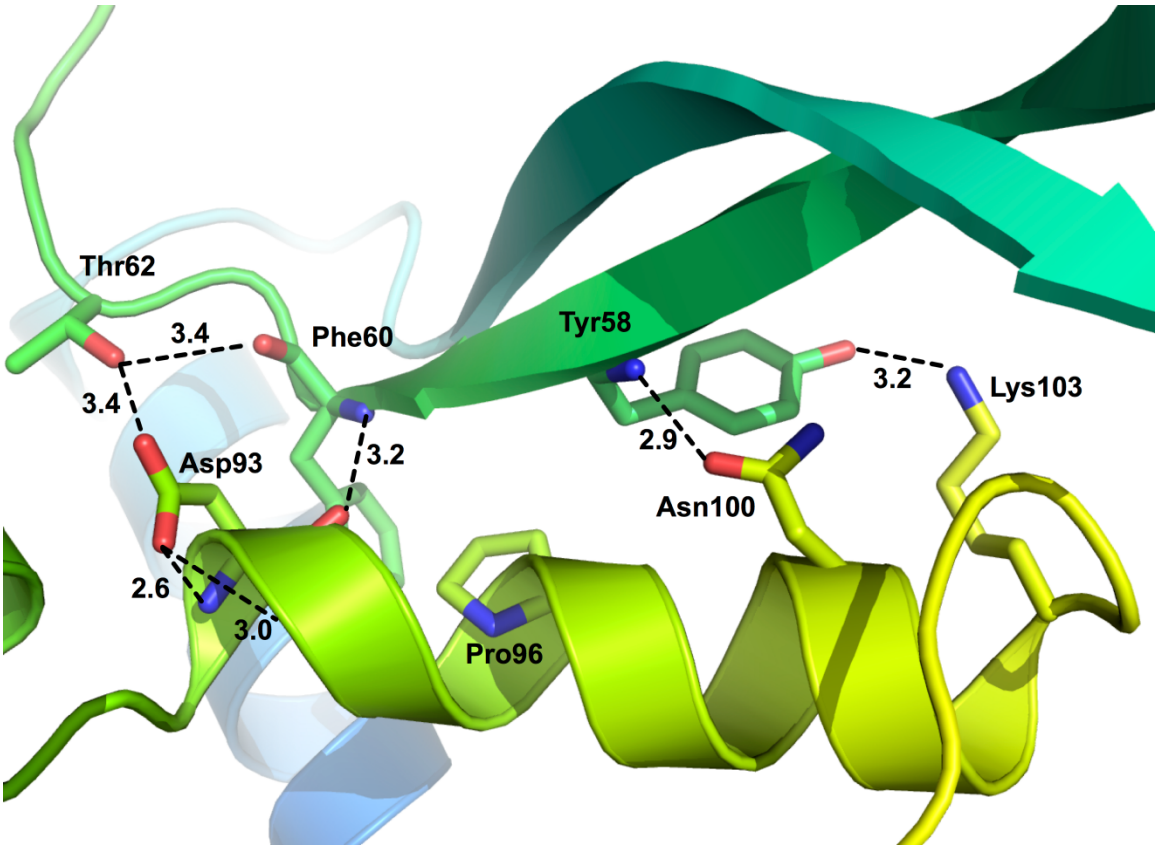

Figure 3.14. The AHABh(alpha-helix-associated beta-hairpin)-motif that defines the superfamily fold is formed by a β-hairpin comprising β-strands 1 and 2 packing against helix D.

Quantitative analysis of helical geometry confirms a kink formed by Pro96 in concert with residues with low helix propensity (89) (Ser91, Asp93, Glu94) distorts the α-helical geometry of the N-terminal end (residues 92-96) of helix D into 3_{10}-helix geometry (figure 3.15; Table 3.2). The 3_{10}-helix geometry is defined by the amide N-H group of each amino acid forming a hydrogen bond with the amide carbonyl group of the amino acid residues three residues prior in sequence

(90, 91). Thus, each turn of the helix formed by the hydrogen bond is comprised of ten atoms, making it a more tightly wound helix than a canonical α-helix. As a result, the 3_{10} geometry is considered relatively unstable in comparison; however, it has been postulated that inherent instability imparts a unique dynamic character to these helices (92). Unsurprisingly, 3_{10}-helices are often found in functionally-dynamic regions of protein structure such as the active-site, voltage switches in ion channels, and heme-iron coordination sites (92, 93). Despite lacking the some of the stabilizing hydrogen bonds canonical to α-helix geometry, the 3_{10}-helix extension of helix D in PglC is stabilized through the presence of the Asp93 at the helical N-cap position (94) to satisfy the positive electrostatic environment of the helix dipole. Additionally, the hydrophobic packing of the two leucine residues (L92 and L95) at the N′ and C_0 positions of the helix add further stability to the secondary structure geometry.

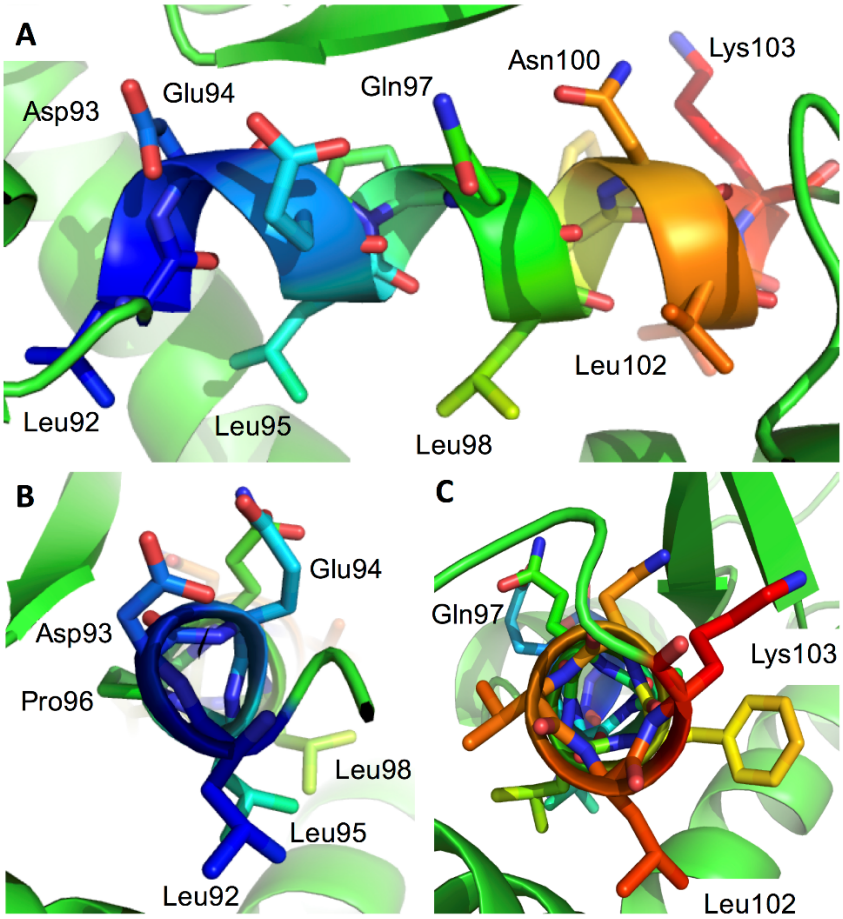

Figure 3.15. Active-site helix D 3_{10}-geometry allows for co-facial positioning of Asp-Glu catalytic dyad. A, Helix D (residues 92- 104) shown in color ramp representation from N-terminus (blue) to C-terminus (red). **B,** Looking down helical axis from N-terminus. N-terminus of helix D adopts 3_{10}–helix geometry. **C,** Looking down helical axis from C-terminus. C-terminus of helix D adopts canonical α-helix geometry.

Table 3.2. Results of helical geometry analysis for helix D.

Helix Partition	Helix Length (residues)	Twist (°)	Residues/ Turn	Unit Height (Å)
Full-length	13	107.2 ± 6.35	3.36 ± 0.20	1.75 ± 0.28
N-terminal	5	112.8 ± 8.08	3.19 ± 0.23	2.06 ± 0.40
C-terminal	9	104.3 ± 5.14	3.45 ± 0.17	1.58 ± 0.14
3_{10} (ideal*)	--	120	3.0	2.0
α (ideal*)	--	105	3.6	1.5

* Ideal helix geometries defined as reported by Barlow and Thornton, 1988 (91).

In PglC, this stablized 3_{10}-helical geometry is both structurally and functionally relevant as it architects proximal positioning of the catalytic-dyad by favoring the observed Asp93 rotomer via interactions with the 3_{10}-helix dipole (Figure 3.14, dashed gray bonds). Participation of one of the Asp93 sidechain oxygen atoms in these helix-capping hydrogen bonds in addition to coordination of the catalytically required Mg^{2+} ion contributes to the nucleophilic reactivity of the non-coordinating Asp oxygen via a disruption of resonance stabilization. This increased nucleophilic reactivity for one oxygen of the Asp sidechain is distinctive of covalent-intermediate enzymatic mechanisms (95). As such, the AHABh-motif is not solely a structural motif in PglC, but also plays a critical role in the catalytic mechanism.

3.4 Conclusions

Independent validations of the topology of PglC with respect to the membrane establish the unique membrane interaction modality inclusive of the reentrant membrane helix (helices A, B) and three co-planar amphipathic helices (C, D, and I). This modality acts to stabilize the minimal functional unit of PglC in the membrane. As previously described, the overall structure and membrane-association mode of PglC is largely divergent from other monotopic proteins of known structure in both domain structure, membrane-association interface, and free energy of transfer to the membrane. Additionally, as there are no known structural homologs, either soluble or membrane-associated, it is plausible that the evolution of this the novel PglC fold was intimately associated with the membrane.

As a result, the extensive membrane-resident volume of PglC together with the membrane in which it is positioned effectively forms a folding core and membrane interaction modality distinct from other structurally characterized monotopic proteins. These monotopic membrane proteins of known structure can be classified by the canonical modes of monotopic protein membrane interaction (Figure 3.2): (A) interactions mediated through an amphipathic helix, (B) hydrophobic loops extending into the membrane, (C) electrostatic interactions

between ions and lipid headgroups, and (D) covalent lipid anchors. From the comparison of membrane interaction interface of PglC versus the other monotopics of known structure, it is clear that the structure of PglC represents a first-in-class example of a new modality in which the membrane interaction interface of PglC is comprised of an extensive hydrophobic membrane-embedded domain formed by the reentrant membrane helix and three amphipathic helices (Figure 3.16).

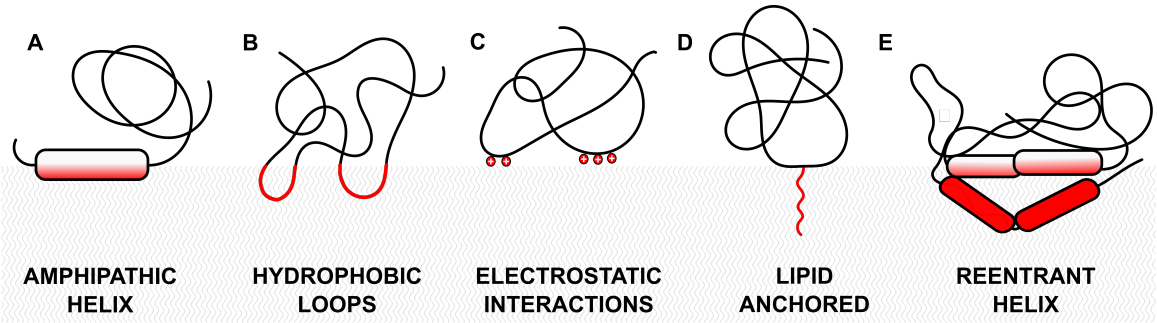

Figure 3.16. Schematic representation of interactions between monotopic membrane proteins and the membrane. A, Interaction via an amphipathic α-helix parallel to the plane of the membrane (prostaglandin H2 synthase – 1CQE); **B,** Interaction via hydrophobic loops (carboxylesterase – 3CN9); **C,** Electrostatic or ionic interactions with lipid head groups (arachidonate 15-lipoxygenase – 2P0M); **D,** membrane association via a covalent fatty acyl, prenyl or glycosylphosphatidyl-inositol post-translational modification (acetylcholinesterase – 1N5M) ; **E,** Reentrant membrane helix (RMH) with C- and N-termini on the same face of the membrane and amphipathic α-helices parallel to the membrane plane (PglC – 5W7L).

Though it is minimal by comparison to other monotopics of known structure, the soluble domain of PglC demonstrates an efficiency with its fold that

is both structurally and functionally significant. Small parsimonious structural motifs not only underpin the critical structural feature of membrane-association, the RMH, they also directly influence active-site geometry through specific positioning of catalytically-important residues (Arg 112 and Asp93). Overall, the unique structure of *C. concisus* PglC represents a novel architecture for membrane interaction and the first structurally characterized member of the monotopic PGT superfamily.

CHAPTER FOUR

Structural underpinnings of PglC-Substrate Interactions

The interpretation of PEG and PO_4^{2-} electron density and phosphate release kinetics presented in this chapter are included in the following accepted article: Ray, L.C., Das, D., Entova, S., Lukose, V., Lynch, A.J., Imperiali, B., and Allen, K.N. Membrane association of monotopic phosphoglycosyl transferase underpins function. *Nature Chemical Biology* (2018). *(in press).*

4.1 Introduction

Although the monotopic and polytopic PGT superfamilies perform the same general PGT reaction, a transfer of a C1-phosphosugar from a UDP-sugar to a polyprenol phosphate embedded in the membrane, they rely on vastly different scaffolds and molecular logic. The structure of PglC, an exemplar of the monotopic PGT superfamily reveals a novel architecture for membrane association and function that is largely divergent from that of the polytopic PGT class. As these two families are so structurally and mechanistically divergent, it follows that the active-site and binding-site geometry should be dissimilar as well.

4.1.1 Catalytic sequence motifs in the PGT superfamily

Previous biochemical and bioinformatics studies of polytopic family members MraY and WecA identified a conserved DDxxD motif across the superfamily essential for catalysis (96). In MraY, the only structurally characterized member of the superfamily, conserved adjacent aspartic acid residues originally proposed to responsible for the coordination of the catalytically required Mg^{2+} ion have been shown to be required for substrate orientation and participation in catalysis (19). This sequence motif is reminiscent of the highly conserved Asp-Glu catalytic-dyad in the monotopic PGT family. However, the

physical geometry of these residues do not follow the same architecture; the co-facial positioning of the adjacent acidic residues enforced by the 3_{10}–helix geometry of helix D in PglC (Figure 3.14) is not mirrored in MraY. The adjacent aspartic acid residues in MraY are part of an α-helix and are thus splayed apart according to canonical α-helix geometry (19). The difference in architecture is subtle (figure 4.1), however the shortest distance between oxygen atoms on the adjacent residues for PglC illustrates the enforced rotomer for PglC Asp93 places the residues 0.5 – 1 Å closer together. As a result, the functional role of the similar motif is likely different in both families. For the monotopic PGTs, the Asp-Glu dyad has been confirmed as the catalytic motif with Asp93 acting as the catalytic nucleophile (25). It should be noted, that trapping of the UDP-sugar intermediate on Glu94 in PglC has been observed, but to a much lesser extent than for Asp93 and most likely is the result of off-path reactions during the experimental protocol (25).

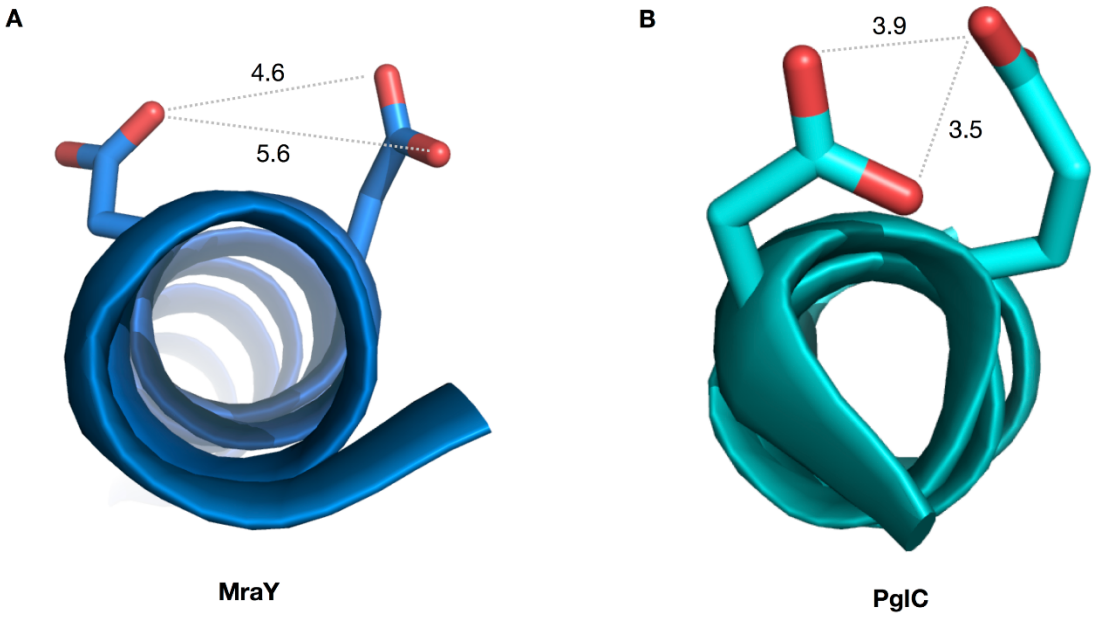

Figure 4.1. Difference in catalytic-dyad architecture for MraY and PglC. A, MraY (blue) residues Asp117 and Asp118 shown as sticks. **B,** PglC (cyan) residues Asp93 and Glu94 shown as sticks. Shortest distances between oxygen atoms on proximal residues (gray dashed lines) given in Å.

In addition to the differences in molecular architecture of the catalytic-dyads and the active-sites as a whole, the polytopic and monotopic PGT superfamiles are mechanistically divergent. MraY and WecA act through a ternary complex reaction mechanism where both substrates, the UDP-sugar and the polyprenyl phosphate, enter the active site at the same time for catalysis (Figure 4.2A) (23, 24). In contrast, that the monotopic PGTs proceed through a covalent-phospho-sugar intermediate enzymatic mechanism which undergoes two-step, ping-pong kinetics. In the first step, the UDP-sugar enters the active site where the

enzyme nucleophile, Asp93, attacks the UDP-sugar to create the phospho-sugar-enzyme intermediate and liberate the mechanistic by-product, UMP. In the second step, the membrane-embedded polyprenol phosphate (pren-P) can enter the active site for catalysis to release the membrane-bound Pren-PP-sugar product (figure 4.2B) (25). As structure begets function, it is expected that the marked difference in enzymatic mechanism should be mirrored by a difference in the molecular architecture of the active site and substrate binding sites between the polytopic and monotopic PGT superfamilies.

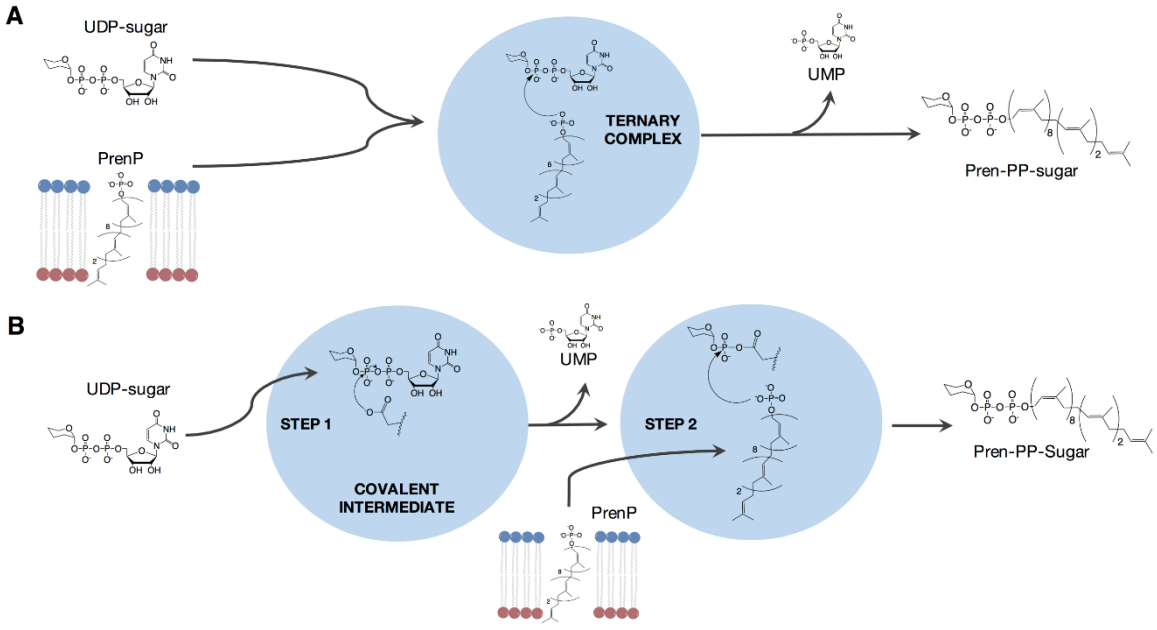

Figure 4.2. Schematic representation of polytopic and monotopic PGT enzymatic mechanisms. Enzyme active site represented by blue circle. **A,** Polytopic PGTs proceed through one-step ternary complex mechanisms. **B,** Monotopic PGTs proceed through two-step covalent-intermediate mechanisms.

4.1.2 Substrate binding sites in PGTs

Biochemical and bioinformatics analyses can predict likely binding sites for the UDP-sugar and Pren-P substrates of PGTs, however these analyses are often met with challenges owing to the complex nature of the substrate and non-conserved binding motifs (97). Sequence conservation analysis for the polytopic superfamily reveals sequence motifs with several conserved histidine residues that are thought to be involved in sugar binding. Nucleoside-inhibitor bound structures of MraY have corroborated the implication of these conserved histidine residues in the positioning of the sugar moiety from localization of the competitive inhibitors in the active site (98). As a consequence of the bi-substrate analog nature of the nucleoside-inhibitor class, binding of the uridine moiety suggests the binding site for the UDP of the native UDP-sugar substrate to be anchored by hydrogen-bonding interactions to the pyrimidine ring and further stabilized through ring-stacking of the uracil moiety.

As no canonical nucleotide-sugar binding site its apparent from bioinformatics or biochemical studies for PglC, one can only postulate as to its architecture. The lack of a conserved canonical-like binding site suggests that UDP-sugar binding in PglC could be driven by electrostatic and steric effects rather than highly specific molecular interactions. Precedent for such a binding

site has been seen in the glycosyltransferase GT-A and GT-B folds. While the GT-B fold possesses a strictly-conserved polar loop biochemically observed to influence sugar specificity adjacent to the putative nucleotide sugar-diphosphate binding-site, exemplified in OleD (99), the GT-fold shows conservation of loops surrounding the DXD-coordinated cation-mediated sugar-nucleotide binding site (100). As a result of the dearth of specific molecular interactions, this type of steric-occlusion-driven binding site is often indiscriminant towards substrates (101). However, it should be noted that promiscuous substrate binding directly counters the observed high-sugar specificity for some PGTs (figure 4.3). In addition to the lack of conserved sugar binding signatures in the superfamilies, the complex nature of the sugar-nucleotides employed in these reactions has made understanding of sugar specificity across PGTs a difficult undertaking with most sugars for biochemical assays of specificity unavailable commercially (15, 97).

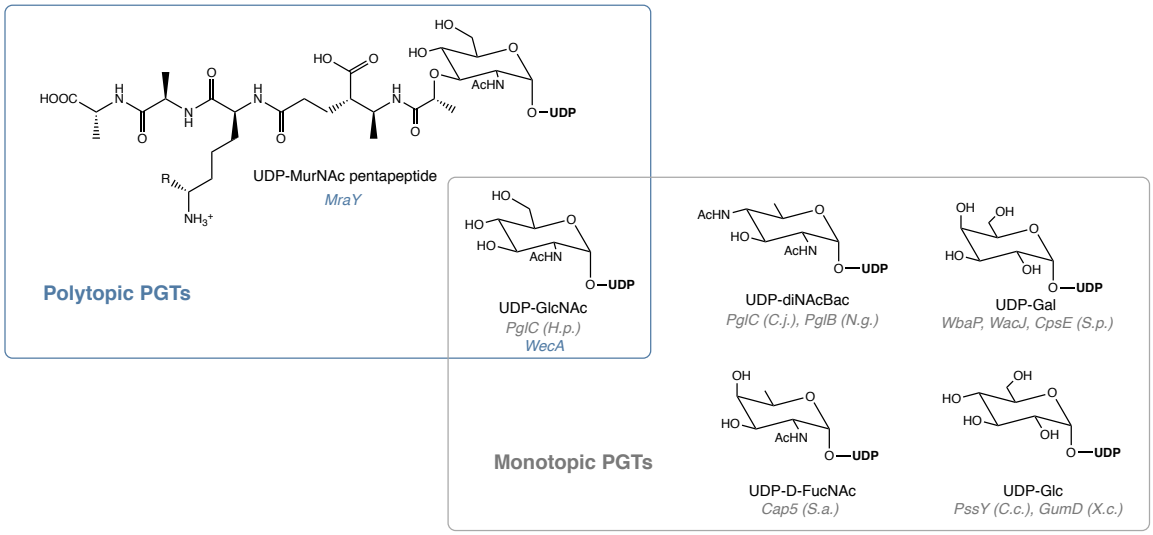

Figure 4.3. UDP-sugar substrates of the Polytopic and Monotopic PGT superfamilies. Polytopic PGT family denoted by a blue box and text, monotopic PGT family denoted by a gray box text. Species names abbreviated as follows: *H.p. = H. pullorum, C.j. = C. jejuni, N.g. = N. gonorroheae, S.a. = S. aureus, S.p. = S. pneumoniae,* and *X.c. = X. campestris.*

The polyprenol phosphate specificity for PGTs is similarly difficult to determine biochemically due to lack of commercially available substrates and difficulty in purification of the native substrates from bacteria. Despite this challenge, it has been shown that members of the polytopic PGT superfamily tend to be less specific for polyprenol chain-length and saturation. Specifically, WecA will accept polyprenols carbon chains made up of 10 to 35 carbons, albeit with lower activity than the native C_{55} substrate (23, 102). MraY also displays a wide range of substrate acceptance and will catalyze the PGT reaction with polyprenols

and dolichols of widely varying chain-length and saturation (102). Structural analysis of MraY affords prediction of the polyprenol phosphate binding site. The structure of MraY reveals a U-shaped hydrophobic cleft extending towards the membrane from the active-site Mg^{2+} cofactor and catalytic aspartic acid (Asp117) (19). Recent crystal structures of MraY in complex with the nucleoside inhibitors tunicamycin and muramycin D2 corroborates the involvement of this hydrophobic cleft in polyprenol phosphate binding. Despite a lack of density for the flexible acyl chain of the inhibitor, the sidechain of Asn172 located within this cleft near the active site was observed to directly interact with the fatty acid amide moieties (98, 103).

In contrast to the broad polyprenol phosphate specificity evident in the polytopic superfamily, monotopic PGTs are characteristically more specific for polyprenol chain length. Both WbaP and PglC only show significant activity with undecaprenol phosphate (104, 105), and are wholly unable to perform the PGT reaction with polyprenol substrates that have fewer than 35 carbons. As there is no lipid binding motif evident biochemically or through bioinformatics, the structure of presented in this dissertation PglC should afford a platform from which to make predictions of a binding site for the UndP substrate.

4.1.3 Implications of active-site and binding site geometry on inhibitor design

Although the inhibitor-bound structures of MraY have provided critical confirmation of the predicted architecture and composition of the active site and natural substrate binding sites in the polytopic superfamily, this knowledge is not easily transferred to the monotopic PGT superfamily. As discussed in Chapter 1, inhibition by the nucleoside inhibitor tunicamycin of the monotopic PGT superfamily, exemplified by PglC, is not seen (20). Clearly, the overall divergence observed between these two families extends to active site geometry and binding sites, as well. The divergent enzymatic mechanisms and molecular structures of the polytopic and monotopic PGT superfamilies provides a basis for understanding the differential sensitivities of these two classes of PGTs to inhibition. Structural understanding of the active site and substrate interactions is needed for the monotopic PGT superfamily to be able to inform targeting of this superfamily of proteins by inhbitors.

4.2 Materials and Methods

4.2.1 *Interpretation of PEG and phosphate electron density in the I57M/Q175M C. concisus PglC electron density map*

The inorganic phosphate (PO_4^{3-}) and the PEG were observed in only one of the two chains in the asymmetric unit. A PO_4^{3-} ion and a Mg^{2+} ion were modeled in Chain A, and a Mg^{2+} ion and ordered water molecule were modeled in Chain B into positive F_o-F_c density contoured to 4 RMSD in COOT. The PEG molecule was modeled into positive F_o-F_c density in Chain A contoured to 2.5 RMSD in COOT. Refinement with the DDM detergent alkyl group (dodecane- Ligand ID:D12) is approximately equivalent to refinement with PEG. However, the absence of any observed density for the maltose disaccharide of the DDM detergent led us to continue the final refinement with the PEG moiety. Notably, the observed density cannot be the undecaprenol moiety of undecaprenol phosphate as the density was observed in all datasets including those collected from crystals which were not co-crystallized with added Und-P.

4.2.2 *Phosphate Release Kinetics Assays*

Phosphate release from UDP by PglC was measured using the Biomol Green reagent (Enzo Life Sciences) for phosphate detection. Purified *C. concisus* PglC (WT, I57M/I87M, I57M/I87M/D93N, and I57M/I87M/E94Q) was diluted to 20

μM in assay buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 5% glycerol, 0.03% DDM) to a final volume of 40 μL. UDP was added at a final concentration of 200 μM in the presence and absence of 5 mM $MgCl_2$. Enzyme and substrate were incubated at room temperature at time points up to 180 minutes in a sealed 96-well microtiter plate. Reactions were quenched by addition of 100 μL Biomol Green reagent to each reaction well. The denatured sample of I57M/I87M PglC was heat-inactivated by incubation at 98 °C for 10 minutes. The plate was incubated for 30 minutes at room temperature for color to develop. Absorbance at λ = 620 nm was measured on a Molecular Devices SpectraMax M5 plate reader. All reactions were performed in triplicate. Inorganic phosphate release was quantified by comparison to a standard curve of KH_2PO_4 in deionized H_2O. Controls with UMP (200 μM), UDP-GlcNAc (80 μM), and catalytically-inactive variants, I57M/I87M/D93N and I57M/I87M/E94Q, (20 μM) were performed to address the mechanism of phosphate release. Separate control phosphate analysis experiments using PglC, UDP, buffer, water, DDM (0.03%) and $MgCl_2$ (5 mM) individually confirmed that none of these reagents contributed to the phosphate measured.

4.2.3 Computational identification of potential substrate binding sites in PglC

To prepare structures for computational docking experiments ligands were removed from Chain B of the model of I57M/Q175M PglC. The 3-D coordinates of UDP-diNAcBac were created by Phenix.eLBOW(106) after input as a SMILES string:

O=C(CCN1C2C(O)C(O)C(COP([O-])(OP(OC3[C@H](NC(C)=O)[C@@H](O)[C@H](NC(C)=O)[C@@H](C)O3)([O-])=O)=O)O2)NC1=O.

The 3D coordinates of UndP (ligand ID: 5TR) were downloaded from the PDB in .cif format. Coordinate files for the structure of PglC and substrates were prepared to for computational docking experiments with AutoDockTools (107) and DockPrep in UCSF Chimera (75) to include polar hydrogen atoms, partial charges, and atom types. PglC-substrate docking was performed using AutoDock Vina (108) implemented through UCSF Chimera with default parameters accepted. The PglC model was used as the receptor for docking of the substrates (UndP and UDP-Bac). Nine binding modes were generated for each PglC-substrate combination. Binding modes with the lowest predicted binding-energy (kcal mol^{-1}) were selected as the likely substrate binding pose.

4.2.4 Crystallographic efforts towards substrate liganded structures

High-throughput crystallization screening at Hauptman-Woodward Institute. Preparation of *C. concisus* PglC constructs (Table 4.1) for high-throughput crystallization screening was performed by Dr. Debasis Das (Imperiali Lab, MIT). All protein samples were provided at a concentration of 6.1 - 6.5 mg/ mL in 50 mM HEPES 7.5, 100 mM NaCl, 5% glycerol, and 0.12% DDM to the HWI HT Crystallization Screening Center (http://hwi.buffalo.edu/high-throughput-crystallization-center/). E93Q PglC containing 1 mM $MgCl_2$ and 520 μM UDP-diNAcBac and WT PglC containing 1 mM $MgCl_2$ and 560 μM UndP were provided for the soluble protein screen. K59R PglC containing 1 mM $MgCl_2$ and 560 μM UDP-diNAcBac was provided for the membrane protein screen. A subset of crystallization conditions resulting in formation of crystalline material based on crystal size, morphology, and crystallization condition components were selected for optimization into higher drop volume, 24-well hanging-drop format for each PglC construct.

Table 4.1. Amino acid sequences of PglC constructs used to attempt to obtain liganded structures. Underline denotes N-terminal methionine residue. Red indicates position of the point mutations.

Construct	Amino acid sequence
WT *	SGSGMYRNFLKRVIDILGALFLLILTSPIIIATAIFIYFKVSRDVIFTQA RPGLNEKIFKIYKFKTMSDERDANGELLPDDQRLGKFGKLIRSLSLDELP QLFNVLKGDMSFIGPRPLLVEYLPIYNETQKHRHDVRPGITGLAQVNGRN AISWEKKFEYDVYYAKNLSFMLDVKIALQTIEKVLKRSGVSKEGQATTEK FNGKN
D93N	SGSGMYRNFLKRVIDILGALFLLILTSPIIIATAIFIYFKVSRDVIFTQA RPGLNEKIFKMYKFKTMSDERDANGELLPDDQRLGKFGKLIRSLSL**N**ELP QLFNVLKGDMSFIGPRPLLVEYLPIYNETQKHRHDVRPGITGLAQVNGRN AISWEKKFEYDVYYAKNLSFMLDVKIALMTIEKVLKRSGVSKEGQATTEK FNGKN
E94Q*	SGSGMYRNFLKRVIDILGALFLLILTSPIIIATAIFIYFKVSRDVIFTQA RPGLNEKIFKMYKFKTMSDERDANGELLPDDQRLGKFGKLMRSLSLD**Q**LP QLFNVLKGDMSFIGPRPLLVEYLPIYNETQKHRHDVRPGITGLAQVNGRN AISWEKKFEYDVYYAKNLSFMLDVKIALQTIEKVLKRSGVSKEGQATTEK FNGKN
K59R*	SGSGMYRNFLKRVIDILGALFLLILTSPIIIATAIFIYFKVSRDVIFTQA RPGLNEKIF**R**MYKFKTMSDERDANGELLPDDQRLGKFGKLMRSLSLDELP QLFNVLKGDMSFIGPRPLLVEYLPIYNETQKHRHDVRPGITGLAQVNGRN AISWEKKFEYDVYYAKNLSFMLDVKIALQTIEKVLKRSGVSKEGQATTEK FNGKN
R112Q	SGSGMYRNFLKRVIDILGALFLLILTSPIIIATAIFIYFKVSRDVIFTQA RPGLNEKIFKMYKFKTMSDERDANGELLPDDQRLGKFGKLMRSLSLDELP QLFNVLKGDMSFIGP**Q**PLLVEYLPIYNETQKHRHDVRPGITGLAQVNGRN AISWEKKFEYDVYYAKNLSFMLDVKIALQTIEKVLKRSGVSKEGQATTEK FNGKN

* Constructs provided to HWI for HT screening.

Crystallization of WT and R112Q C. concicus PglC. Crystallization conditions identified by the HWI high-throughput screen for WT PglC co-crystallized with 1 mM $MgCl_2$ and 560 μM UndP (0.2 M calcium chloride dehydrate pH 5.1, and 20%

PEG 3350 (w/v) reproduced crystals in 24-well hanging-drop crystallization experiment format. Optimization of these crystallization conditions by varying concentration of calcium chloride and precipitant (PEG 3350) produced crystals of WT and R112Q variant *C. concisus* PglC which diffracted to sufficiently high resolution for data collection (3.0-2.2 Å). The specific crystals used for data collection were grown from crystallization experiments set up at 17 °C with crystallization reagents purchased from Hampton Research used after temperature equilibration. All co-crystallization experiments were performed via an 30 minute incubation step on ice and subsequent centrifugation step at 14,000 rpm for 10 minutes at 4 ° C . WT PglC (280 μM) co-crystallized with 1 mM $MgCl_2$ and 560 μM UndP (2 equivalents) produced optimized crystals in 0.225 M $CaCl_2$ pH 5.0 and 21% PEG 3350 and a drop ratio of 2 μL of PglC : 3 μL of well solution. Crystals appeared within 3 days and were fully grown in 10 days. R112Q PglC (280 μM) co-crystallized with 1mM $MgCl_2$ and 560 μM UndP (2 equivalents) produced optimized crystals in similar conditions (0.2 M $CaCl_2$ pH 5.0 and 23% PEG 3350) and a drop ratio of 2 μL of PglC : 2 μL of well solution. R112Q variant crystals appeared in 2 days and were fully grown in 5 days. All crystals used in data collection were flash-cooled in liquid nitrogen for transport to the beamlines without additional cryoprotection.

Phasing and refinement of R112Q and WT C. concisus PglC structures. A WT PglC dataset was collected from a crystal diffracting to 2.19 Å at beamline 24-ID-E (12665 eV, 100 K) at the Advanced Photon Source (Chicago, IL). Additionally, a dataset from R112Q PglC was collected from a crystal diffracting to 2.33 Å at beamline 24-ID-E (12665 eV, 100 K) the Advanced Photon Source (Chicago, IL). Indexing of the datasets collected for both WT and R112Q PglC found the space group P 3_2 2 1 and unit cell dimensions of a = b = 71.198 Å, c = 189.64 Å; α=β= 90°, γ = 120° for WT, and a = b = 71.136 Å, c = 189.539 Å; α=β= 90°, γ = 120° for R112Q PglC. Matthews coefficient analyses for these datasets suggests 2 copies in the asymmetric unit (ASU). As the unit-cell dimensions and space group assignments identified by indexing were isomorphous to the I57M/Q175M *C. concisus* PglC dataset used to solve the molecular structure presented in Chapter 2 of this dissertation, isomorphous replacement was used phase the WT and R112Q datasets. The resulting models of R112Q and WT PglC are undergoing continued refinement with Phenix.refine (43). Current data collection and refinement statistics are summarized in Table 4.2.

Table 4.2. Data collection and refinement statistics for WT and R112Q Variant PglC.

	WT (Dataset ID: C2x5)	**R112Q variant** (Dataset ID: C2x14)
Data collection		
Beamline	APS 24-ID-E	APS 24-ID-E
Wavelength (Å)	1.00	0.97918
Resolution range (Å)	58.638 - 2.19 (2.26- 2.19)	63.180 - 2.330 (2.41 - 2.33)
Space group	P 3_2 2 1	P 3_2 2 1
Unit Cell (Å)	a = b = 71.198; c=189.64	a = b = 71.41; c = 189.54
Total Reflections	29630 (2235)	24382 (2396)
Unique reflections	29277	24451
Multiplicity	4.8 (4.3)	4.3 (4.5)
Completeness (%)	82.76 (76.47)	98.39 (99.46)
Mean I/sigma(I)	13.7 (1.5)	12.3 (1.0)
Wilson B-factor	53.09	66.43
Rmerge	0.047 (1.77)	0.054 (1.22)
Rmeas	0.059 (2.254)	0.07 (1.53)
$CC_{1/2}$	0.99 (0.790)	0.99 (0.397)
Refinement		
Refl. used in refinement	25460 (2504)	24382 (2396)
Relfections used for Rfree	3754 (340)	3370 (381)
R_{work}	0.2942 (0.3696)	0.2352 (0.3227)
R_{free}	0.3340 (0.3872)	0.2905 (0.3494)
CC_{work}	0.845	0.899
CC_{free}	0.808	0.81
Number of non-hydrogen atoms	3015	3017
Protein residues	370	370
RMS(bonds)	0.005	0.003
RMS(angles)	0.98	0.62
Ramachandran favored (%)	93.72	95.9
Ramachandran outliers (%)	1.37	1.37
Rotomer outliers (%)	0.00	1.52
Clashscore	13.46	7.47
Average B-factor	79.2	92.8

4.2.5 Structural analyses of sequence conservation presented by Lukose, et al. (2015).

Sequence conservation across the monotopic PGT superfamily was mapped onto the structure of I57M/Q175M *C. concisus* PglC in UCSF Chimera (75) with the Multalign Viewer tool. The multiple sequence alignment reported by Lukose, *et al*. (2015) (17) was loaded into Chimera and the sequence for *C. jejuni* (Uniprot ID: O86156) was associated with the structure for mapping. Sequence conservation was rendered on the structure in color ramp on the structure with blue denoting high sequence conservation and red denoting low sequence conservation. Additional assessment of sequence conservation for the PglC-like subfamily in comparison to all three monotopic PGT subfamilies was achieved by rendering residues with 70% sequence conservation in blue.

4.2.6 Structural analysis of large unit cell

A dataset of WT SeMet derivatized PglC was collected to 3.01 Å from a crystal grown in 100 mM BisTris pH 6.0, 350 mM $MgCl_2$, and 26% PEG 3350 at 17 °C with solutions purchased from Hampton Research and temperature equilibrated. Crystallization was carried out following the crystallization protocol for WT SeMet PglC described in section 2.2.1. The crystal used for data collection grew over 21 days and was cryocooled in liquid nitrogen for transportation to the

beamline without any additional cryoprotection. The 3.01 Å WT SeMet dataset was collected at Brookhaven National Lab (BNL) beamline 17-ID-1 during beamline commissioning experiments at a wavelength of 1.0 Å and a temperature of 100 K. The dataset was indexed into P 3_1 2 1 with unit-cell dimensions of $a = b =$ 142.82 Å, $c = 192.563$ Å; $\alpha = \beta = 90°$, $\gamma = 120°$. Indexing, integration, and scaling were performed in HKL2000 (109). Matthews coefficient analyses suggest 8 copies of PglC in the ASU. Data collection and refinement statistics are summarized in Table 4.3. Crystal packing in the large unit cell is highly similar to packing in the small unit cell with only subtle rearrangements resulting in the doubling of unit cell axes *a* and *b.* The WT SeMet dataset is undergoing continued refinement with Phenix.refine (43).

Table 4.3. Data collection and refinement statistics for WT SeMet Large-Unit-Cell Dataset

	WT SeMet (Dataset ID: J2)
Data collection	
Beamline	BNL NSLS-II 17-ID-1 (AMX)
Wavelength (Å)	1.00
Resoution range (Å)	47.737 – 3.013 (3.121- 3.01)
Space group	P 3_1 2 1
Unit Cell (Å)	a = b = 142.82; c = 192.563
Total Reflections	75337
Unique reflections	38779 (3544)
Multiplicity	1.9 (2.0)
Completeness (%)	82.76 (76.47)
Mean I/sigma(I)	14.2 (1.0)
Wilson B-factor	84.21
Rmerge	0.074 (0.656)
Rmeas	0.105 (0.928)
$CC_{1/2}$	0.98 (0.553)
Refinement	
Reflections used in refinement	38417 (3426)
Relfections used for Rfree	3754 (340)
R_{work}	0.2708 (0.3585)
R_{free}	0.2833 (0.3651)
CC_{work}	0.811
CC_{free}	0.764
Number of non-hydrogen atoms	12113
Protein residues	1480
RMS(bonds)	0.002
RMS(angles)	0.613
Ramachandran favored (%)	95.77
Ramachandran outliers (%)	0.34
Rotomer outliers (%)	0.34
Clashscore	7.92
Average b-factor	115.38

4.3 Results and Discussion

The structurally-efficient scaffold of PglC imparts membrane-association and function through small structural motifs that establish the RMH and directly influence active-site geometry through specific arrangement of catalytically-important residues. The distinct co-facial positioning of the essential catalytic-dyad residues (Asp93 and Glu94), underpinned by the 3_{10}-helix of the AHABh-motif (Figure 3.15) is further enforced by through the coordination of the Mg^{2+} co-factor by Asp93 (figure 4.4). Notably, coordination of Mg^{2+} by absolutely conserved adjacent acidic residues has likewise been predicted biochemically in the monotopic PGT superfamily members WecA and WbaP (105, 110). In the crystal structure of PglC, the Mg^{2+} co-factor is further coordinated by the main-chain carbonyl-oxygen conserved of Arg88 and the phosphate ion observed in crystallization.

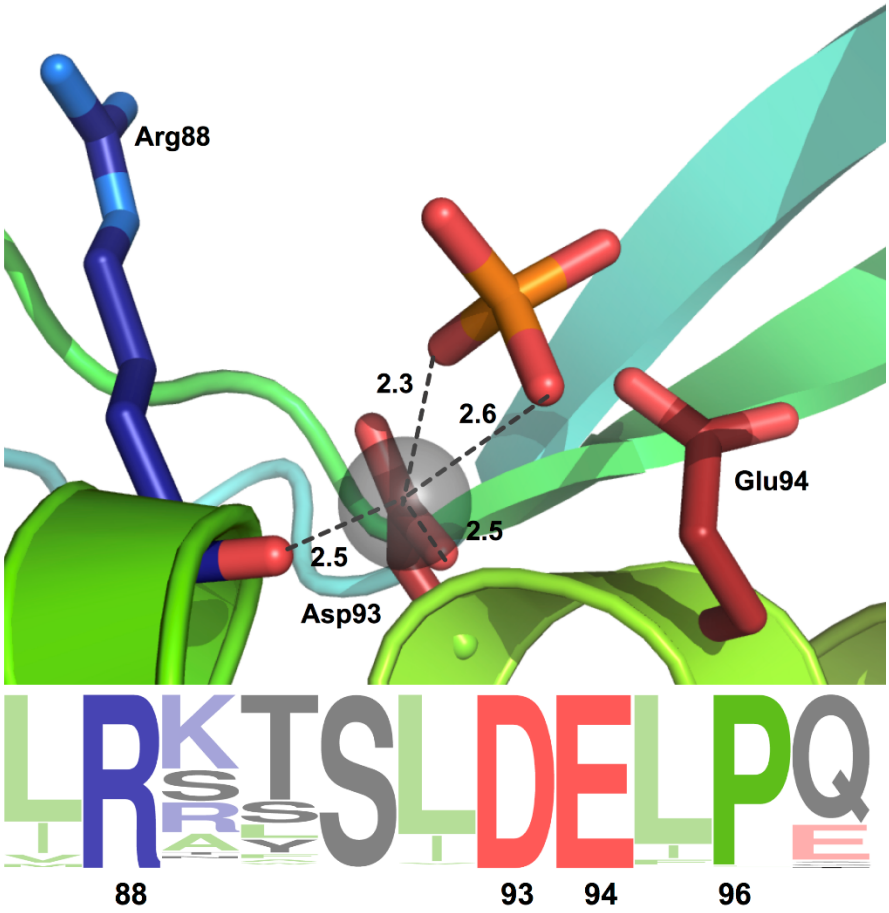

Figure 4.4. Active-site architecture. Depiction of the active-site showing the conserved Asp-Glu dyad with Mg^{2+} and phosphate ligands and sequence logo. Phosphate was not added in the protein preparation or crystallization conditions, but resulted from hydrolysis of exogenously-added 5-iodo-UDP during crystallization.

Surprisingly, catalytically-important interactions are not observed for the side chain of Arg88, thus the absolute conservation of this residue across the superfamily must arise from a separate mechanism of evolutionary conservation. One such mechanism could be the evolutionary-pressure for conservation of the

electrostatic environment of the active-site (Figure 4.5). Mapping of the electrostatic charge distribution calculated by the Poisson-Boltzmann equation (via APBS, see methods in section 3.2.4) (73) onto the molecular structure of PglC reveals numerous basic residues forming a positive-electrostatic funnel poised for binding and orienting the negatively-charged phosphate-rich UndP and UDP-diNAcBac substrates in the active-site (Figure 4.4).

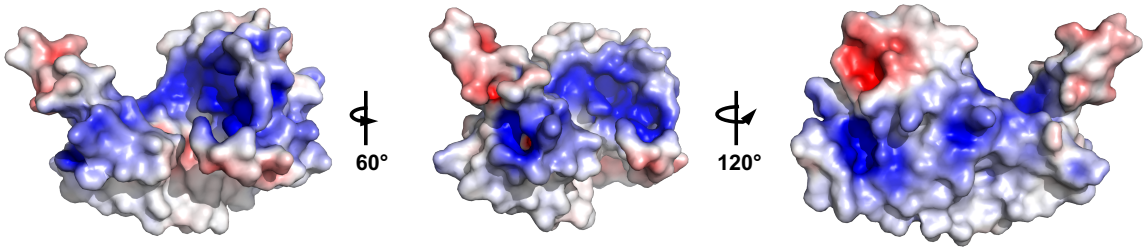

Figure 4.5. Electrostatic surface representations of PglC colored by charge from red (-5 kT/e) to blue (+5 kT/e).

4.3.1 PEG andPhosphate binding suggest location of UndP and phosphate-binding subsites

As discussed briefly in chapter 2, the structure of I57M/Q175M PglC was refined with a PEG molecule occupying an active-site cleft and an inorganic phosphate molecule proximal to the catalytic Asp-Glu dyad (Figure 4.6). Presence of a molecule of PEG bound within the soluble and membrane-accessible cleft identifies the putative UndP binding subsite. The geometry ascribed by the

putative-binding subsite would allow for the membrane-embedded UndP to access the catalytic dyad, and therefore the phospho-sugar intermediate, without leaving the membrane. Whereas the PEG molecule is more polar and flexible than the isoprenoid moiety of the UndP, the observation within crystal structures of PEG molecules bound at hydrophobic-substrate binding-sites localized at the membrane interface is not without precedent (111). Additionally, analysis of the hydrophobicity of the residues lining the putative UndP-binding cleft illustrates an amphipathic directionality to the site. Residues below the predicted membrane plane exhibit greater hydrophobicity than those above the membrane plane suggesting that the hydrophobic isoprenoid and phosphate moieties of the UndP substrate could remain in a chemically favorable environment within the binding cleft simultaneously.

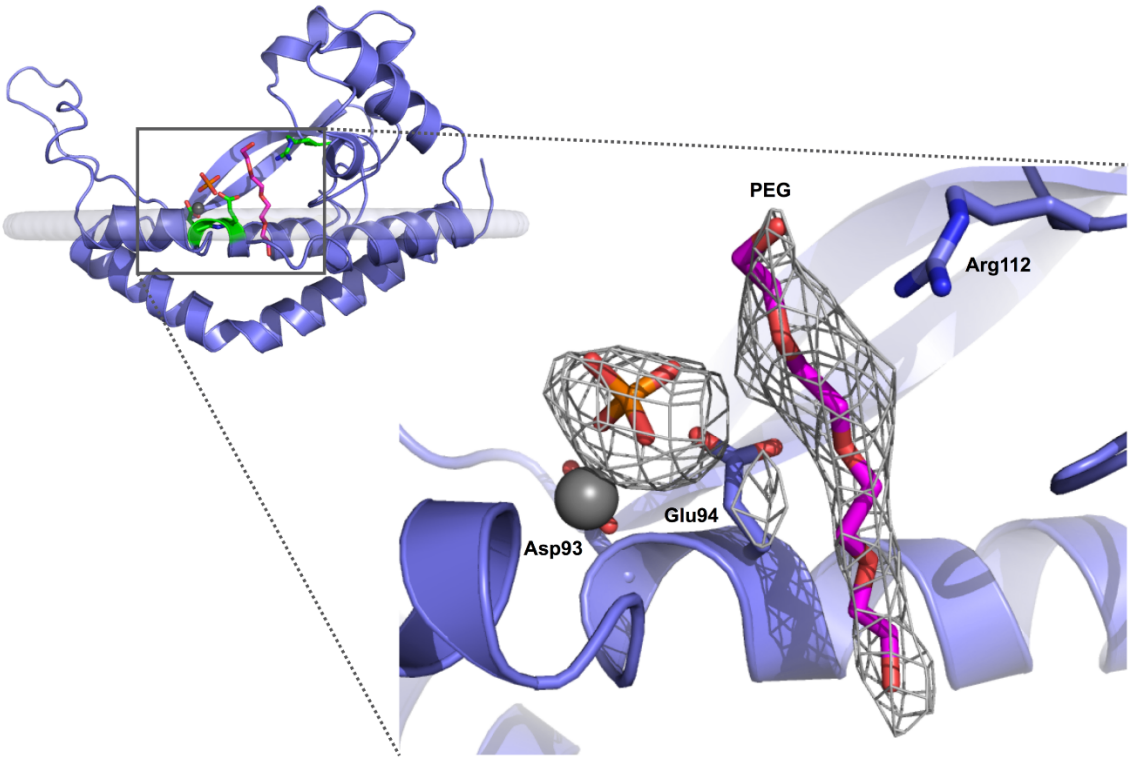

Figure 4.6. Polyethylene glycol (PEG) position in the PglC binding site identifies the putative Pren-P binding site. PEG (pink), Arg112, Glu94, Asp93 (green), and phosphate (orange) are shown as sticks and Mg^{2+} is depicted as a grey sphere. The hypothetical membrane surface is shown as a transparent grey surface. Inset, F_o-F_c simulated annealing omit map contoured to 3σ shown as gray wire mesh. PEG molecule 305 (chain A) shown in pink sticks. Phosphate molecule 302 (chain A) shown in orange sticks.

The presence of phosphate within the structure also suggests a putative substrate binding-subsite for PglC. Phosphate informs on the putative phosphate-binding subsite of UDP-diNAcBac or phosphor-sugar intermediate moiety. Notably, phosphate was not added in the protein preparation or crystallization

conditions, therefore its observation within the crystal structure resulted from hydrolysis of exogenously-added 5-iodo-UDP during crystallization. As UDP is not the native substrate for PglC, evaluation of enzymatic phosphate release form PglC constructs used in crystallization incubated with UDP were carried performed using the Biomol Green assay (Figure 4.7). A measured rate of free phosphate liberation from UDP of 1.76 nM s^{-1} confirms PglC is able to turn over UDP only in the presence of the catalytically-required Mg^{2+} despite the large difference in charge between the di-anionic UDP and native UDP-diNAcBac substrates. Additionally, enzymatic phosphate release from UDP by variants of PglC which are catalytically-inactive against the native substrate (I57M/I87M/D93N and I57M/I87M/D93N) suggests the mechanism of phosphate release in the enzyme differs from that of the native substrate. Further support for a differing mechanism of phosphate-release for non-native substrate is provided by measurement of a relatively similar rate of UDP turn over in a competition assay with catalytically-active PglC and a significantly less catalytically-active substrate-mimic (25), UDP-N-acetylglucosamine (UDP-GlcNAc).

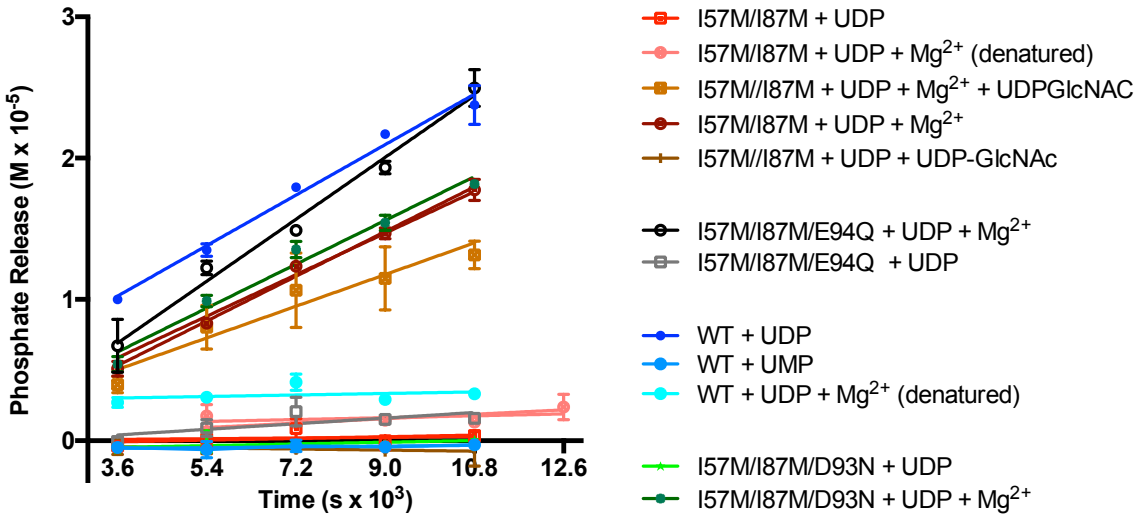

Figure 4.7. Time course of phosphate release in the reaction of 20 μM PglC with 200 μM UDP in the presence and absence of 5 mM $MgCl_2$. Assays were carried out in triplicate. Error bars represent mean ± standard deviation. Reaction with I57M/I87M PglC at 20 μM with 200 μM UDP and 5 mM $MgCl_2$ gave a rate of phosphate release from UDP of 1.76 nM s^{-1} (brick red, circles). Reaction withI57M/I87M/E94Q PglC at 20 μM with 200 μM UDP and 5 mM $MgCl_2$ gave a rate of phosphate release from UDP of 2.44 nM s^{-1} (black, open circles). Reaction with I57M/I87M/D93N PglC at 20 μM with 200 μM UDP and 5 mM $MgCl_2$ gave a rate of phosphate release from UDP of 1.72 nM s^{-1} (green, circles). Reaction with I57M/I87M PglC at 20 μM with 200 μM UDP and 80 µM UDP-GlcNAc and 5 mM $MgCl_2$ gave a rate of phosphate release from UDP of 1.25 nM s^{-1} (green, circles) (orange, squares). Reaction with WT PglC at 20 μM with 200 μM UDP and 5 mM $MgCl_2$ gave a rate of phosphate release from UDP of 1.99 nM s^{-1} (blue, circles). Separate control phosphate analysis experiments using PglC, UDP, buffer, water, DDM (0.03%) and $MgCl_2$ (5 mM) individually confirmed that none of these reagents contributed to the phosphate measured.

4.3.2 *Computational docking of UDP-diNAcBac and UndP corroborates predicted substrate binding sites for PglC*

In the absence of experimentally-determined PglC-substrate complex structures, computational docking of the native substrates UDP-diNAcBac and UndP into the enzyme active-site performed using AutoDock Vina (108) corroborates the positioning of the binding subsites indicated by the phosphate and PEG moieties in the crystal structure. The enthalpies of the binding poses calculated for UndP were all similar ranging from -6.7 kcal mol^{-1} to -6.1 kcal mol^{-1}. The highest scoring binding pose for UndP (Figure 4.8A) positions the UndP isoprenoid moiety into the hydrophobic, putatively-membrane embedded, cleft and the UndP phosphate moiety within close proximity to both the catalytic dyad and Arg112. Arg112 has been shown to be important in positioning of substrates through mutagenesis studies (17). Similarly, the binding pose of UDP-diNAcBac chosen as the best pose places the di-phosphate linker between the nucleoside and sugar groups in close proximity to both catalytically-important centers of functionality. It should be noted that the binding pose with the lowest binding enthalpy (-8.0 kcal mol^{-1}) for UDP-diNAcBac did not dock the molecule within close proximity to both Arg112 and the Asp-Glu catalytic dyad. Thus, a binding pose with a slightly less favorable binding energy (-7.4 kcal mol^{-1}) that positioned

the substrate within close proximity to both catalytically-important groups was chosen as the best, most-likely binding pose (Figure 4.8B).

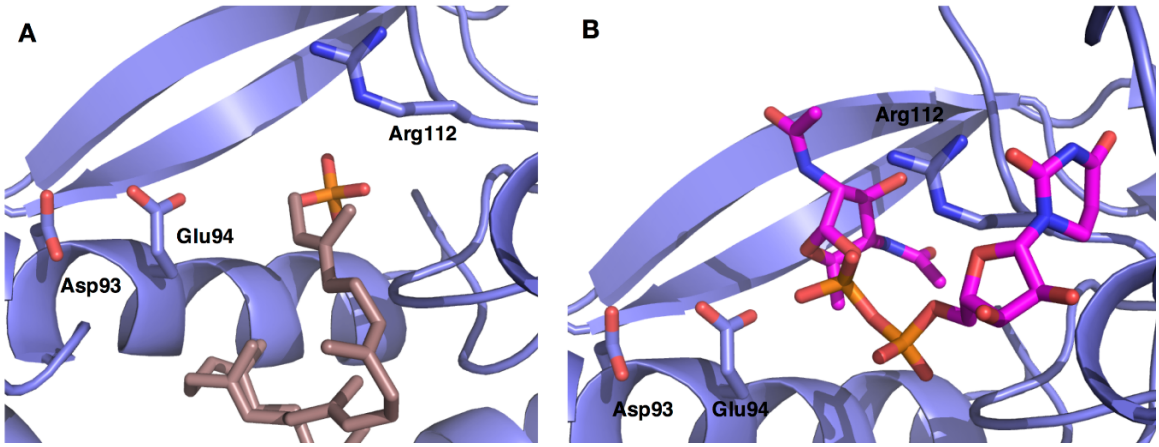

Figure 4.8. Substrate binding poses computed by AutoDock Vina in proximity to the catalytic Asp-Glu dyad and Arg112 (blue sticks). A, Binding pose predicted for UndP (rose sticks). **B,** Binding pose predicted for UDP-diNAcBac (magenta sticks).

In the context of the covalent phospho-sugar intermediate mechanism described previously for *C. concisus* PglC (25), which dictates that the UndP and UDP-diNAcBac substrates cannot be resident within the enzyme active site at the same time, the binding sites for the β-phosphate of the UDP-diNAcBac and the UndP are predicted to be coincident. Consistent with this prediction, overlaying the chosen best docked poses for the UDP-diNAcBac and UndP within the active-

site clearly illustrates coincidence in the binding sites for the UDP-sugar, β-phosphate, and UndP phosphoryl moiety (Figure 4.9).

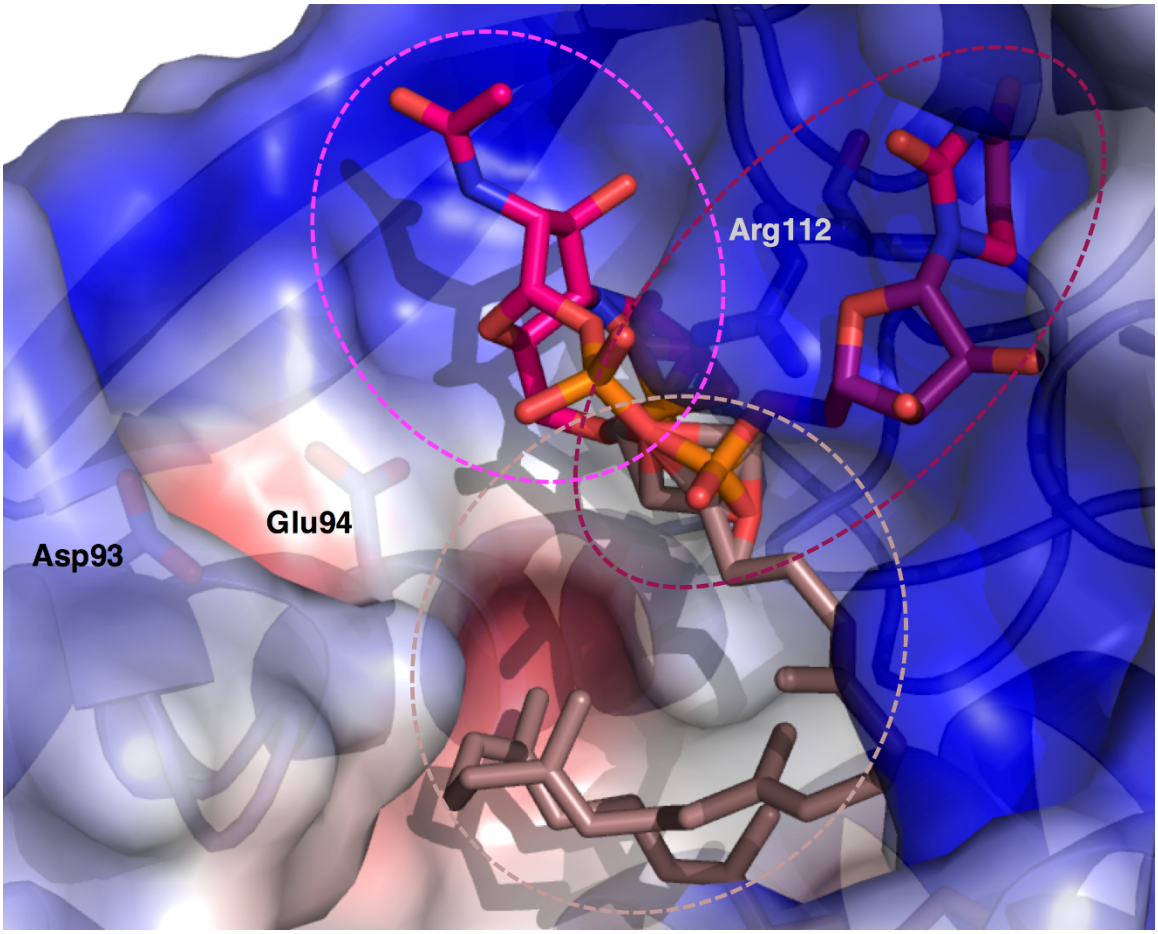

Figure 4.9. Overlay of docked poses reveals coincident binding sites for UndP and and UDP-diNAcBac. UDP binding subsite indicated with purple dashed circle, UndP binding subsite indicated by rose dashed circle, and sugar binding subsite indicated by magenta dashed circle. UDP-diNAcBac displayed with uracil nucleoside represented by purple sticks, phosphoryl moiety represented by orange sticks, and diNAcBac moiety represented by magenta sticks. UndP displayed with isoprenoid tail represented by rose sticks and β-phosphate moiety represented by orange sticks.

The functional residues of PglC identified through conservation and mutagenesis by Lukose *et al.* (2015) (17) map to the structure in coordination with the predicted-binding sites. Pro24, a determinant of the kink in the RMH, was implicated in interactions with UndP, this follows with the proposed binding-site as Pro24 lies within the hydrophobic cleft in which the UndP isoprenoid moiety is docked. Arg112, identified to interact with phosphate groups by mutagenesis plays a critical function in positioning both the phosphoryl groups from UndP and UDP-diNAcBac. Notably, it is possible that guanidinium functional-group of Arg112 is important for orienting the uridine ring of UDP-diNAcBac within the activesite. However, those interactions were not observed in any of the top 10 binding-poses identified by AutoDockVina. Subtle ligand-induced binding interactions may be missed as docking of substrates is performed with a relatively static molecule of PglC. Given the current best docked pose from these docking experiments, Asp93 and Glu94 comprising the catalytic dyad are well positioned in the PglC active site for catalysis given the predicted substrate-binding sites.

4.3.3 *Predicted UndP and UDP-binding subsites are conserved across the monotopic PGT superfamily*

All monotopic PGTs follow a general reaction scheme where the PGT catalyzes the transfer of a C1-phosphsugar to from a UDP-sugar to UndP. Notably, the utilization of UDP and UndP is universally conserved across the superfamily; however, the identity of the sugar moiety accepted by the individual members varies within the subfamilies and superfamily with some PglC-like subfamily members (*C. concisus, C. jejuni)* utilizing diNAcBac as their sugar substrate (20) and others utilizing GlcNAc (*H. pullorum) (112)*. Consistent with this, analysis of amino-acid conservation across the monotopic PGT subfamilies and superfamily as a whole illustrates that the predicted UDP and UndP binding subsites display a high degree of conservation in primary structure (Figure 4.10). Similarly, as expected, the residues composing the predicted sugar-binding subsite are markedly less conserved to accommodate differing sugar specificities across the superfamily.

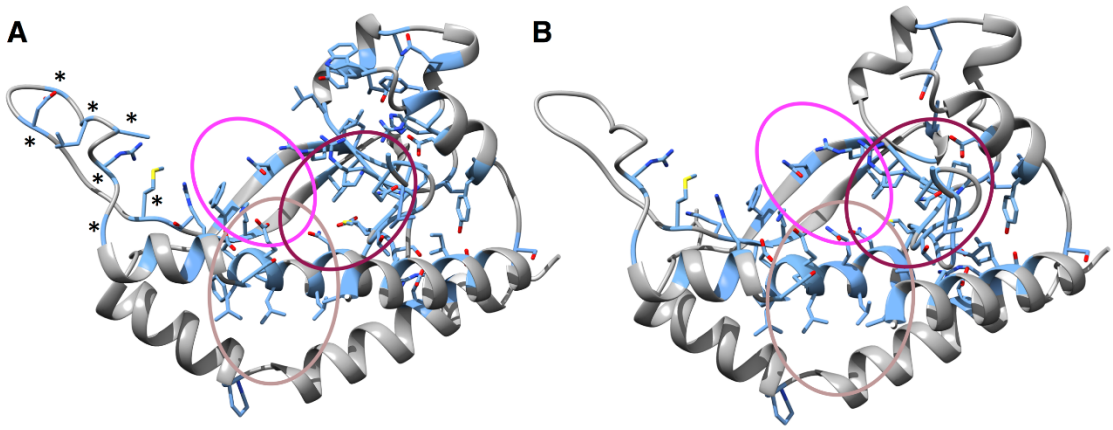

Figure 4.10. Predicted UDP and UndP binding-site is well conserved across the three PGT families. A, sequence conservation for the PglC-like family mapped to the structure with residues with ≥70% conservation shown as blue sticks. Starred residues denote residues of the extended loop with 100% sequence conservation. **B,** Sequence conservation for all three monotopic PGT subfamilies. Residues with ≥ 70% conservation shown as blue sticks.

Unexpectedly, the residues within the extended loop connecting β-strand 2 with helix C (aa 61- 81) are well conserved in the PglC-like subfamily (Figure 4.10A, starred). This loop, termed the ANGEL loop based on the sequence forming the turn (61-KMTMSDERDANGELLPDDQRLG-81), was predicted by protein structure prediction algorithms to fold down over the active site. The conformation of this loop was the main difference observed between the predicted models and the model determined crystallographically; however, it should be noted that the crystal contacts in the ASU position the ANGEL loops of the two

protomers proximal to each other (Figure 2.8) which results in the pinning of this loop in both protomers into the observed open conformation.

4.3.4 Active site and binding site locations are energetically advantageous

The location of the active site and proposed substrate-binding sites positions all reaction components at the membrane interface. This tactical placement of the active site obviates the need for energetically-costly translocation of the membrane-embedded substrate and product out of, and back to the membrane. In contrast to the molecular logic of PglC, the polyprenol phosphate glycosyl transferase, GtrB, responsible for the biosynthesis of a membrane bound pholyprenol monophosphosugar, requires a translocation of approximately 15 Å of the lipophilic substrate from the membrane to the soluble glycosyltransferase domain where catalysis takes place and back again upon product-release (113). Additionally, the molecular logic of active sites located far from the membrane interface despite membrane-bound substrates and products is re-capitulated in many monotopic enzymes (114) (figure 4.11). In these enzymes, hydrophobic channels allow for the translocation of membrane-bound substrates out of the membrane to the active site (115). As a result, the molecular logic of PglC, as an exemplar for the monotopic PGT superfamily, accomplishes catalysis with

hydrophobic membrane-bound substrates and products while obviating the need energetically-costly translocations of hydrophobic substrates through tactical placement of the active site at the membrane interface.

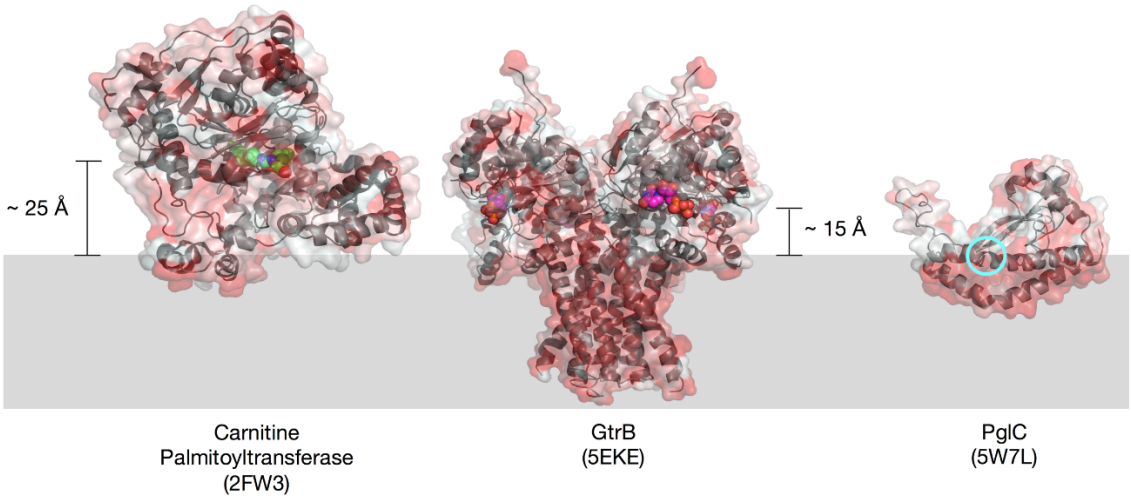

Figure 4.11. Location of enzyme active sites in relation to the membrane. Left, Carnitine palmitoyltransferase shown with its ligand (ID: BUI, green spheres) marking the active site position. **Middle,** GtrB with UDP (pink spheres) marking the active site. **Right,** PglC with cyan circle marking the catalytic-dyad. Protein surface colored by normalized consensus hydrophobicity from hydrophobic (red) to hydrophilic (white). Membrane represented by gray box. Location of membrane plane calculated with PPM server (http://opm.phar.umich.edu/server.php).

Overall, the locations of the catalytic dyad, Mg^{2+} cofactor, and substrate-binding sites that are defined through structural analysis and computational docking afford a mechanistic scheme for PglC catalysis consistent with formation of a covalent phosphosugar intermediate (25) (figure 4.12).

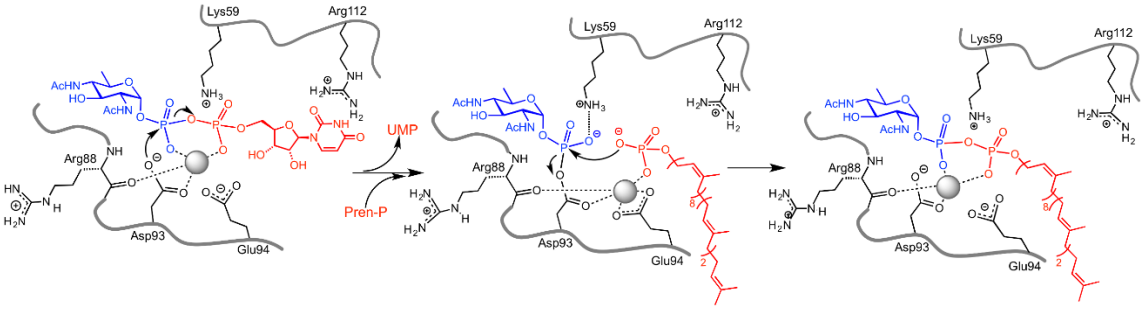

Figure 4.12. The structure of PglC provides insight into the molecular basis of the chemical mechanism of monotopic PGTs. Mg^{2+} is shown as gray sphere. Figure from Ray, L.C., Das, D., Entova, S., Lukose, V., Lynch, A.J., Imperiali, B., and Allen, K.N. Membrane association of monotopic phosphoglycosyl transferase underpins function. *Nature Chemical Biology* (2018). *(accepted).*

4.3.5 Conformational change could be required for tight substrate binding

Computational and bioinformatics analysis affords the ability to predict and understand the determinants of substrate binding and therefore mechanism in PglC; however, experimental structural data is needed to corroborate predictions and postulations made from such analyses. Attempts to obtain crystal structures of PglC in complex with its substrates with have been largely unsuccessful. As result, closure of the ANGEL loop over the active site, as discussed previously, has been hypothesized to be required for substrate binding. This hypothesis is based on three main observations: the divergence of loop position between energy-minimized structure-prediction models and the

crystallographically-determined model (Figure 2.8B), absolute conservation of many residues in the ANGEL loop (Figure 4.9A, starred), and lastly, the variable loop position correlating to presence or absence of observed positive electron-density in the active sites of promoters in the large unit-cell dataset ASUs.

Although the low resolution (3.01 Å) of the large unit-cell dataset, and likely low-occupancy binding of substrates, precludes modeling of specific ligands into positive F_o-F_c density observed in the active sites of specific promoters, correlations between presence or absence of un-modeled density and loop conformation can be made. When the eight copies of PglC in the ASU of large-unit cell datasets are overlaid, the ANGEL loop is evident as the main region of structural divergence between the chains (Figure 4.13A). In chains G and H, comparatively large amounts of positive density are present. Chains C, D, E, and F all have minimal positive density present. Lastly, chains A and B, have moderate amounts of density present (figure 4.13B). Analysis of ANGEL loop conformation in the context of active-site occupancy shows chains with more positive density appear to be in a more closed conformation than those with little or no density present. Chains occupying the same position in the crystal lattice dimer do not do always occupy the same ANGEL loop position, thus the difference in loop conformation is likely due to differential binding at the active-site. Furthermore, mapping of

residue B-factors for each overlaid chain to understand the structural basis for this divergence reveals that the loops occupy either one of two relatively low-B-factor conformations, open or closed, with higher-B-factor loops occupying positions between these two conformations (Figure 4.13C).

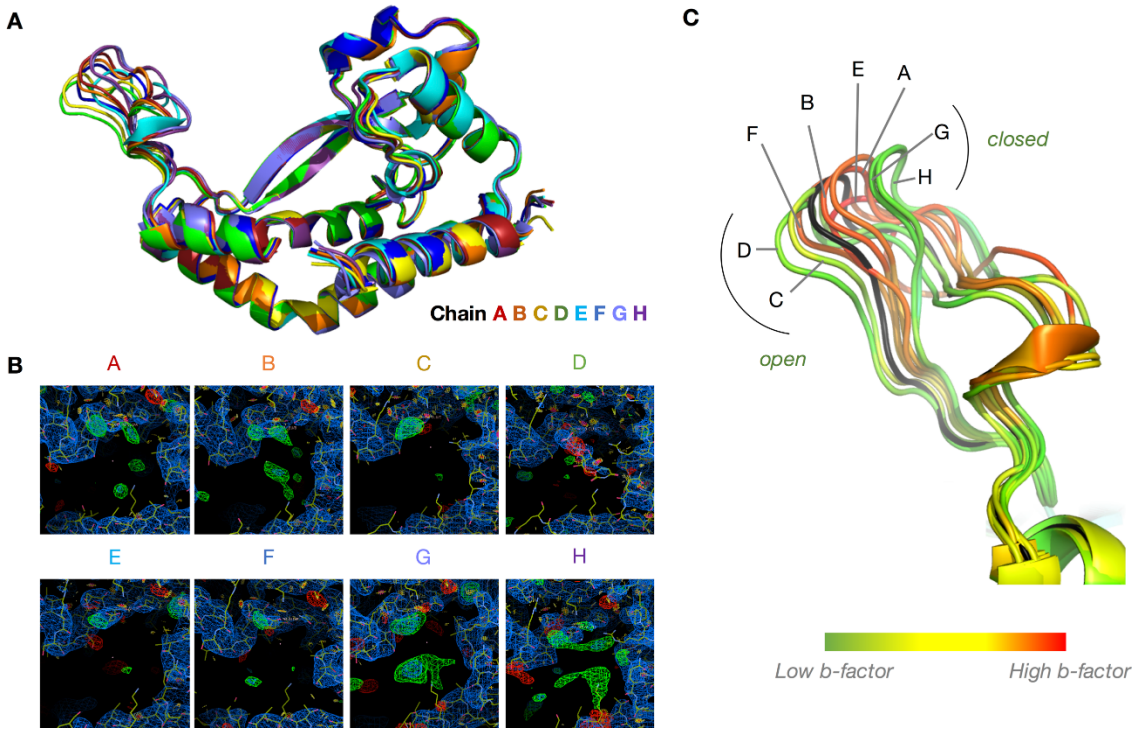

Figure 4.13. ANGEL loop can move to a more "closed" position upon occupancy of binding sites in the activesite. A, Overlay of the 8 chains in the WT SeMet large unit-cell dataset (see Table 4.3). Chains colored by rainbow from chain A (red) to chain H (violet). **B,** 2FO-FC (blue, contoured to 1.5 σ) and FO-FC (green and red, contoured to 3 σ) electron density observed in the active-site of each chain (yellow) show positive density in chains G and H, no density in chains C and D, and small areas of positive density in chains E, F, A, and B. **C,** Low b-factor "open" and "closed" conformations are seen for chains with differential occupancy of the

activesite. B-factors for each residue mapped to the ANGEL loop. Final refined model for I57M/Q175M PglC shown in black.

4.3.6 High-throughput screening and optimization of novel PglC crystallization conditions

All crystals of PglC grown in similar conditions crystallize in the same space groups, P 3_1 2 1 for the large unit cell or P 3_2 2 1 for the small unit cell, and without interpretable ligand-density despite co-crystallization with different substrates, substrate-mimics, or additives; it is unlikely given the current crystallization conditions will afford a ligand-bound structure of PglC. As a result, three PglC constructs were provided for high-throughput crystallization screening at HWI to identify crystallization conditions which result in differing crystal forms (116). Each construct was co-crystallized with substrate in an attempt to find conditions which promoted PglC-substrate complex formation. To avoid turn-over of substrate during crystallization, UDP-diNAcBac was co-crystallized with a catalytically-inactive E94Q and K59R variants. As UndP requires turnover of UDP-diNAcBac to be a viable substrate, WT PglC can be co-crystallized with UndP without enzymatic turnover. The top crystallization hits resulting in the formation of large crystals from HT screening at HWI are summarized in table 4.4. From the initial hit conditions identified, crystallization conditions which were chemically

different from the current crystallization conditions (BisTris pH 6, $MgCl_2$, and PEG 3350) and produced crystals with different crystal morphology than the regular hexagonal and trapezoidal prisms and plates typically seen were selected for optimization (Figure 4.14).

Table 4.4. Initial crystallizaiton conditions selected from HWI HT crystallization screening. Conditions selected for optimization denoted by box.

Construct	Co-crystallizaiton	HWI Cocktail	Salt	Buffer	Precipitant
E94Q	+ 560 μM UDP-	17_C0124	5% (w/v) Ethylammonium nitrate (C2H8N2O3)	0.09 M HEPES pH: 6.8	PEG 3350 27% (w/v)
	+ 5 mM MgCl2	17_C0437	0.1 M Sodium molybdate dihydrate (Na2MoO4*2H2O)	0.1 M Bis-Tris Propane pH: 7	PEG 8000 20% (w/v)
		17_C0453	0.1 M Lithium sulfate monohydrate (Li2SO4*H2O)	0.1 M TAPS pH: 9	PEG 8000 20% (w/v)
		17_C0554	0.1 M Ammonium phosphate-monobasic (NH4H2PO4)	0.1 M Bis-Tris Propane pH: 7	PEG 4000 20% (w/v)
		17_C0600	0.1 M Sodium molybdate dihydrate (Na2MoO4*2H2O)	0.1 M CAPS pH: 10	PEG 4000 20% (w/v)
		17_C0989		0.1 M Sodium citrate tribasic dihydrate pH: 5.5	PEG 3,350 18% w/v
		17_C1001		0.1 M Sodium citrate tribasic dihydrate pH: 5.5	PEG 8,000 16% w/v
		17_C1007		0.1 M Imidazole pH: 7.0	PEG 20,000 12% w/v
		17_C1186		0.2 M Potassium sulfate pH: 6.8	PEG 3350 20%(w/v)
		17_C1216		8%(v/v) Tacsimate pH: 7	PEG 3350 20%(w/v)
		17_C1218		8%(v/v) Tacsimate pH: 8	PEG 3350 20%(w/v)
		17_C1220		0.2 M Succinic acid pH: 7	PEG 3350 20%(w/v)
		17_C1430	0.20M Potassium Sodium Tartrate tetrahydrate		PEG 3350 20%(w/v)
WT	+ 560 μM UndP	17_C0462	0.1 M Manganese sulfate monohydrate	0.1 M Sodium Acetate pH: 5	PEG 8000 20% (w/v)
	+ 5 mM MgCl2	17_C1159		0.2 M Calcium chloride dihydrate pH: 5.1	PEG 3350 20%(w/v)
		17_C1220		0.2 M Succinic acid pH: 7	PEG 3350 20%(w/v)
K59R	+ 560 μM UDP-	M8_C0084	0.4 M Potassium Chloride (KCl)	0.1 M Sodium Citrate pH: 5.6	PEG 400 50% (v/v)
	+ 5 mM MgCl2	M8_C0095	0.4 M Sodium Chloride (NaCl)	0.1 M Sodium Citrate pH: 5.6	PEG 400 40% (v/v)
		M8_C0242	0.2 M Ammonium Sulfate ((NH4)2SO4)	0.1 M TRIS pH: 8.5	PEG 400 40% (v/v)
		M8_C0243	0.2 M Ammonium Sulfate ((NH4)2SO4)	0.1 M TRIS pH: 8.5	PEG 400 50% (v/v)
		M8_C0266	0.1 M Potassium Chloride (KCl)	0.1 M TRIS pH: 8.5	PEG 400 30% (v/v)
		M8_C0308	0.05 M Sodium Chloride (NaCl)	0.1 M Sodium Citrate pH: 5.6	PEG 4000 20% (w/v)
		M8_C0431	0.3 M Magnesium Chloride hexahydrate (MgCl2 *	0.1 M HEPES pH: 7.0	PEG 4000 20% (w/v)
		M8_C0469	0.1 M Sodium Chloride (NaCl)	0.1 M HEPES pH: 7.0	PEG 4000 16% (w/v)
		M8_C0665	0.4 M Magnesium Chloride hexahydrate (MgCl2 *	0.1 M HEPES pH: 7.0	PEG 2000 30% (w/v)
		M8_C1101	0.4 M Magnesium Chloride hexahydrate (MgCl2 *	0.1 M TRIS pH: 8.5	PEG 6000 20% (w/v)
		M8_C1259	0.1 M Lithium Sulfate monohydrate (Li2SO4 * H2O)	0.1 M HEPES pH: 7.0	PEG 3350 20% (w/v)
		M8_C1369		0.1 M Sodium citrate tribasic dihydrate	Jeffamine, pH 7.0
		M8_C1420	18% v/v 2-Propanol	0.1 M Sodium citrate tribasic dihydrate	PEG 4,000 20% w/v

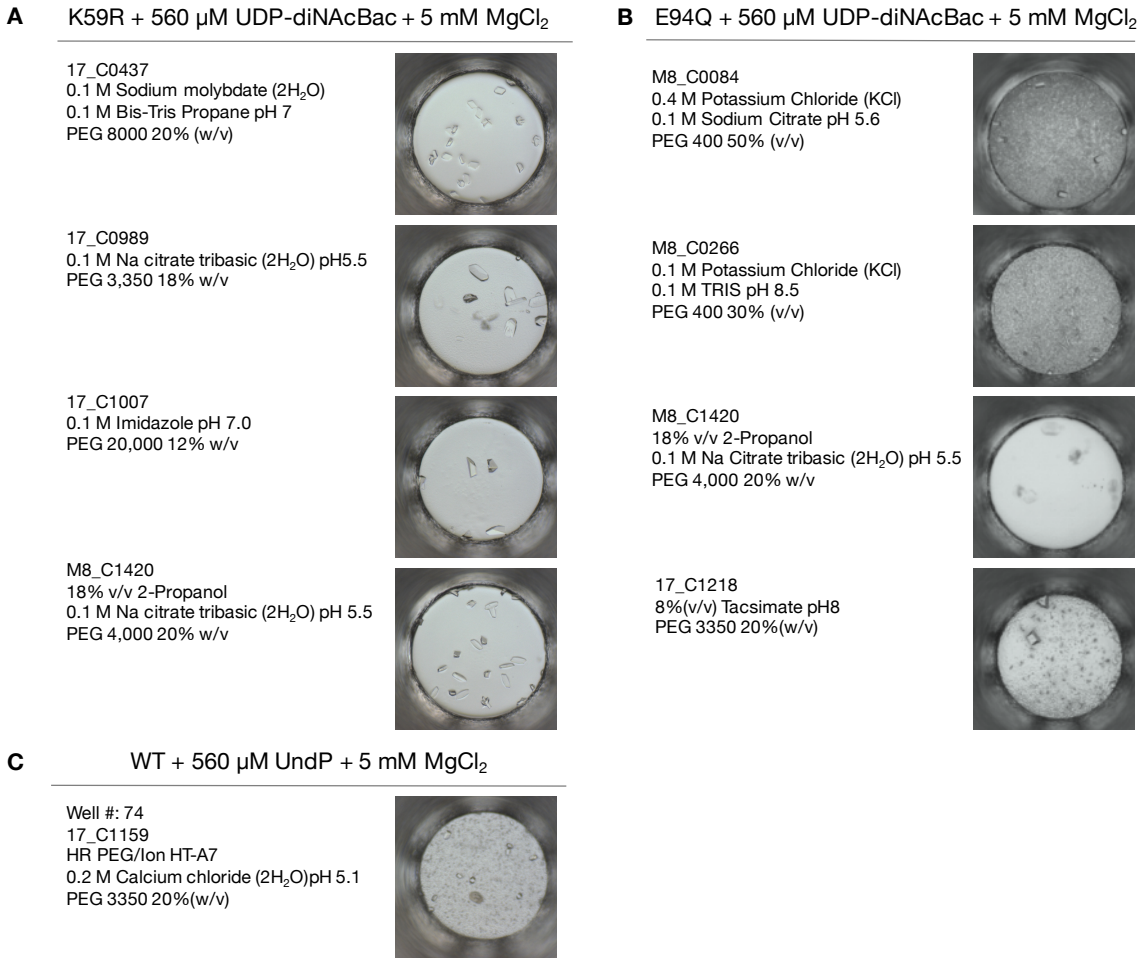

Figure 4.14. Selected crystal conditions for optimization from HWI HT screening. **A,** Crystal hits of K59R PglC co-crystallized with 560 µM UDP-diNAcBac and 5 mM $MgCl_2$. **B,** Crystal hits of E94Q PglC co-crystallized with 560 µM UDP-diNAcBac and 5 mM $MgCl_2$. **C,** Crystal hits of WT PglC co-crystallized with 560 µM UndP and 5 mM $MgCl_2$.

The first round of optimization of the selected crystallization conditions in 24-well hanging-drop format resulted in replication of high-quality crystals only from the $CaCl_2$ and PEG 3350 condition for WT co-crystallized with UndP and $MgCl_2$ and a new variant, R112Q, co-crystallized with UndP and $MgCl_2$.

Unfortunately, this crystallization condition is most similar to the previous crystallization conditions. Conversely, only thin crystals of E94Q PglC co-crystallized with UDP-diNAcBac were obtained from grid-screens around the selected tacsimate and PEG 3350 conditions, while other conditions selected for optimization did not reproduce crystals. The crystals of WT PglC co-crystallized with UndP produced diffraction to 3.0 - 2.19 Å. Similarly, screening of R112Q crystals for diffraction revealed crystals grown in 0.2 M $CaCl_2$ and 23% PEG 3350 produced diffraction to 2.6 - 2.33 Å. Unsurprisingly, the thin crystals grown in 8% tacsimate and 22-26% PEG 3350 only produced reflections to 8-10 Å resolution.

Indexing of the datasets from crystals of WT and R112Q grown in $CaCl_2$ and PEG3350 found similar unit cell dimensions and identical space groups found for crystals grown in the previously used BisTris pH 6, $MgCl_2$, and PEG 3350 condition. As a result, the goal of obtaining a new crystal form and therefore novel crystal packing was unsuccessful; however, data collection at APS beamline 24-ID-E from crystals of WT and R112Q PglC resulted in the determination of two new higher-resolution structures of PglC. Collection of data from a crystal of WT PglC co-crystallized with 560 μM UndP and 5 mM $MgCl_2$ grown in 0.25M $CaCl_2$ and 21% PEG 3350 allowed the determination of the structure to a resolution of 2.19 Å via isomorphous replacement with the structure of I57M/Q175M PglC.

Refinement of this model is still ongoing, with a current R_{work}/R_{free} of 0.2942/0.3340. Likewise, collection of a dataset from a crystal of R112Q co-crystallized with 560 μM UndP and 5 mM $MgCl_2$ grown in a similar condition (0.2 M $CaCl_2$ and 23% PEG 3350) allowed the structure of R112Q PglC to be solved to 2.33 Å. Refinement of this model is ongoing with a current R_{work}/R_{free} of 0.2352/0.3494 (Table 4.2). Although at higher resolution, clear density for UndP was not observed in either dataset, and is therefore not able to be modeled. Notably, in both datasets, density similar to that observed for PEG in the I57M/Q175M model of PglC was observed. All three molecular structures are highly similar with an RMSD = 0.312 Å for WT aligned to I57M/Q175M, and RMSD = 0.250 Å for R112Q aligned to I57M/Q175M.

4.4 Conclusions and Future Directions

The structural underpinnings of substrate interaction and mechanism revealed by structural, biochemical, and computational analyses of the model of I57M/Q175M *C. concisus* PglC support the colvalent phospho-sugar intermediate mechanism proposed for PglC, and the monotopic PGT superfamily overall.

Unfortunately, the current crystallization condition appears non-ideal for crystallization of PglC-substrate complexed structures, as the ANGEL loop hypothesized to close over the active-site upon substrate binding is pinned open

in the crystal lattice produced in this condition. High-throughput screening of novel crystallization conditions and constructs with substrate produced crystals in many novel conditions. While the crystallization conditions selected for the initial round of optimization produced two new higher-resolution molecular models of PglC, WT to 2.19 Å and R112Q variant to 2.33 Å, both structures were un-liganded and packed in the same crystal lattice observed previously. Notably, during this first round of optimization, many of the crystal conditions selected from the HT screen produced poor-diffracting crystals or no crystals. Therefore, current efforts towards developing crystallization conditions to yield PglC-substrate-complexed structures are focused on continued optimization of HT screening hits in crystallization conditions disparate from the current condition in chemical space. In addition, crystallization screening with additives known to mediate crystal contacts, 18-crown-6 ether and Anderson-Evans polyoxotungstate, is ongoing to promote the formation of differing crystal morphologies in an effort to capture the ANGEL loop in a closed conformation (117, 118). In parallel, construction of sequence similarity and genome-neighborhood networks for the PglC-like subfamily of monotopic PGTs will be undertaken to understand sequence determinants of sugar specificity within the subfamily in the absence of experimentally determined UDP-diNAcBac and UndP liganded structures.

Overall, further molecular understanding of the structural determinants of binding of substrates to PglC garnered through the work provided in this chapter and on-going experiments will provide a structural platform from which to address development of anti-virulence agents against the entirety of the monotopic PGT superfamily.

CHAPTER FIVE

Towards Understanding PglC in a Native-Like Lipid Environment

5.1 Introduction

Biophysical characterization and insight into the structure of PglC, primarily via small-angle x-ray Scattering (SAXS) with the use of lipid-bilayer nanodiscs (ND), will allow for an understanding of the structure and function of PglC within an environment which mimics the physiological membrane in which PglC is embedded. Although the highest achievable resolution for many biophysical studies makes discerning atomic features impossible, these low-resolution structural studies are highly complementary to higher resolution methods. Despite the comparably low-resolution, SAXS and analytical ultracentrifugation (AUC) experiments are advantageous in that they characterize proteins in solution in aqueous buffers, and thus, provide an understanding of the protein in a more native-like environment, in direct contrast to the non-physiological highly ordered crystal lattice. The study of membrane proteins via biophysical methods is marked by challenges due to inherent insolubility and propensity for aggregation (31, 33). This challenge is often abrogated through the use of detergents; however, a protein solubilized by detergent is not particularly physiologically relevant (119-121). To move towards studying these proteins in the most native-like environments feasible, nanodiscs (ND) coupled SAXS present as an optimal tool for low-resolution structural studies of membrane proteins.

5.1.1 Lipid-bilayer nanodiscs as a tool to solubilize membrane proteins

Nanodiscs are discrete native-like, discoidal membrane systems with defined composition and predictable nano-scale dimensions (8 – 16 nm). Structurally, ND are supramolecular self-assembled entities composed of a phospholipid bilayer constrained by two copies of a membrane scaffold protein (MSP) engineered from an Apolipoprotein A1 scaffold (122, 123). When these components solubilized in detergent are mixed in precise ratios, ND of defined size and composition will assemble spontaneously upon the removal of the detergent (figure 5.1). The unique, well defined structure imparts ideal physical properties for the utility of ND as a tool in biophysical studies. Most notably, as the hydrophobic environments required by the phospholipid bilayer core is satisfied by the amphipathic helices of the MSP scaffold, ND are soluble in aqueous buffers and therefore can provide stabilization of integral membrane proteins, like PglC, without added detergents.

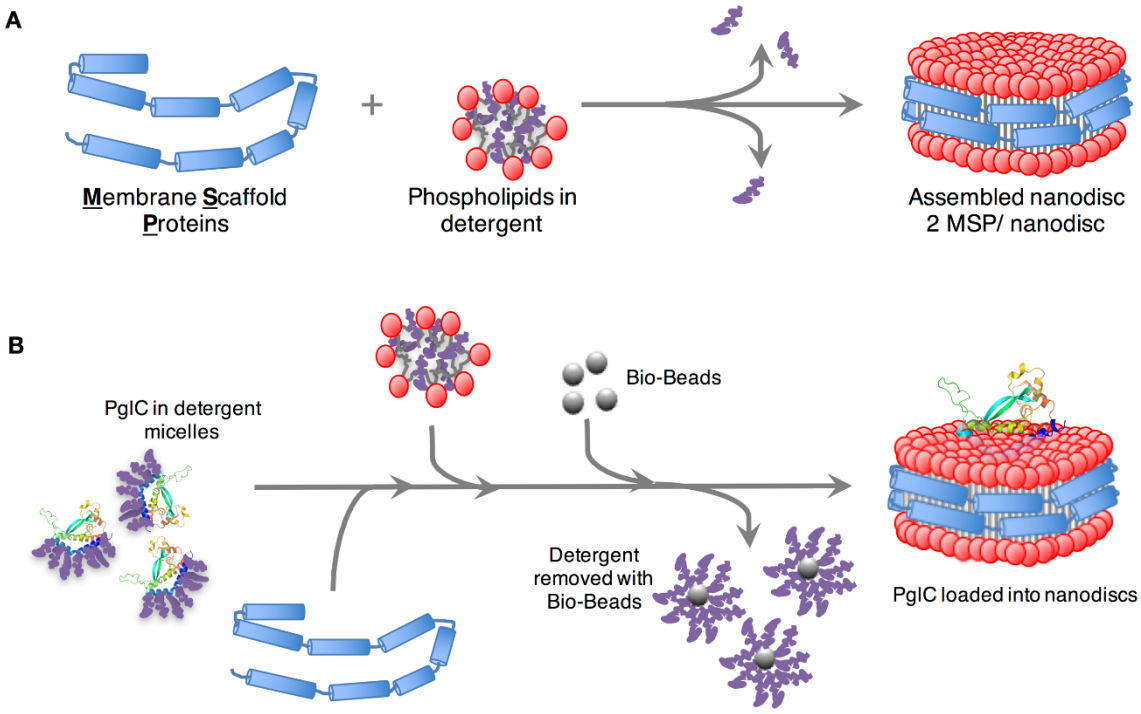

Figure 5.1. Spontaneous assembly (A) and loading (B) of lipid-bilayer nanodiscs.

5.1.2 ND enable studies of PglC by SAXS

With the aid of ND, we can pursue low-resolution structural information for PglC in a native-like lipid environment via SAXS. In general, SAXS is a contrast-based method which measures small-angle scattering of macromolecules that arise from the difference in the average electron density between macromolecules and and that of the surrounding solvent. The simultaneous observation of randomly oriented of macromolecule in solution in SAXS experiments results measurement of isotropic, spherically-averaged scattering

intensities from the macromolecule. In comparison to x-ray crystallography in which scattering intensity amplitudes are measured according to the discrete three-dimensional function, $I(q_{hkl})$, corresponding to the reciprocal lattice, the spherically-averaged intensity measured in SAXS decays as a function of momentum transfer. Ultimately, the distribution of scattering intensity amplitudes into reflections afforded by the crystal lattices leads to measurement of high-resolution data in crystallography, while the spherically-averaged SAXS scattering patterns can only be measured to low resolution. Scattering patterns collected during SAXS experiments are rotationally-averaged to give one-dimensional scattering intensity profile, I(q) as function of the momentum transfer, $q = 4\pi\sin\theta/\lambda$, where λ is the incident wavelength in the experiment, and 2θ represents the scattering angle (figure 5.2). Ultimately, the shape of the scattering curve obtained from SAXS experiments is determined based on the fluctuations of scattering-length density in comparison to the solvent. As such, if the scattering length density of the solvent, principally H_2O in aqueous buffers, is equivalent to to that of the sample, the contrast, or excess scattering length density described by $\Delta\rho = \rho_{sample} - \rho_{solvent}$, is negligible resulting in no scattering from the sample. Contrast presents a unique challenge to the study of membrane proteins solubilized in ND, as the scattering length densities of lipids (ρ_{lipid} = 300 e nm^{-3}) is

roughly equivalent to that of water (ρ_{H2O} = 334 e nm^{-3}), resulting in no scattering measured from the lipid component of the ND (124, 125).

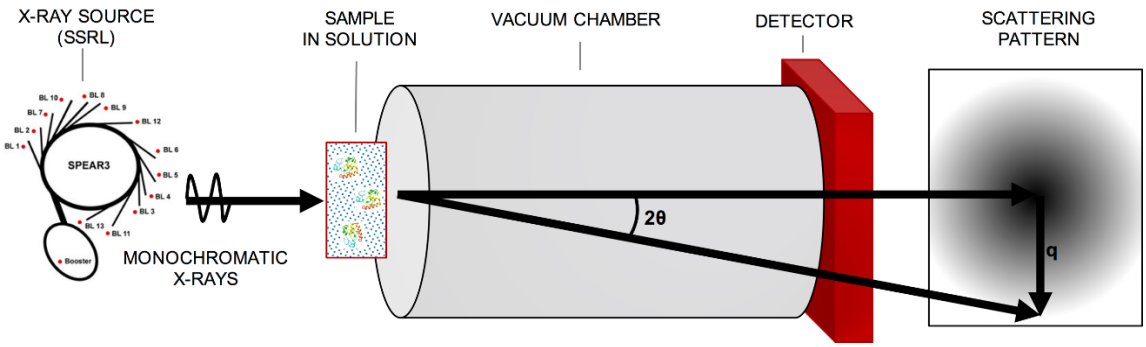

Figure 5.2. Schematic representation of a SAXS experiment.

The power of SAXS lies in the subtraction of the solvent scattering signal measured from a matched-buffer blank from the scattering signal from the macromolecule in solution. Thus, the resultant scattering intensity profile is that of the macromolecule alone. Transformations of the macromolecular SAXS intensity profile allow interpretation of structural parameters relating to the size and overall shape of the macromolecule in solution. One such parameter is radius of gyration, R_g, which provides a measure of the average weighted center-of-mass distances of the electron density for the particle, thus giving an approximation of particle size.

The Guinier approximation (126),

$$I(q) = I(0)e^{\frac{-q^2 R_g{}^2}{3}} \tag{5.1}$$

allows calculation of R_g and I(0), the intensity of scattering at the zero-angle ($2\theta = 0$), directly from the scattering function, I(q). However, this approximation is only valid at very small scattering agles such that $q < 1.3/R_g$. Linearization of this approximation yields the a format for the standard equation of a line; thus, linear fitting to the plot of ln[I(q)] vs q^2 gives direct measurement of R_g from the slope of the line and I(o) from the y-intercept (127). Measurement of I(0) is critical for determination of the molecular mass of the macromolecule in solution via SAXS. Similarly, the scattering intensity data can be represented in real-space through indirect Fourier transformation resulting in the pairwise distance distribution function, *P(r)*, given by,

$$I(q) = 4\pi \int_0^{D_{max}} P(r) \cdot \frac{\sin(qr)}{qr} dr \tag{5.2}$$

where, the scattering intensity I(q) is described by an integral limited to the maximal dimension or interparticle vector, D_{max} (124).

5.1.3 *SAXS as a method for determining molecular mass of a sample in solution*

The subtractive nature of SAXS can be leveraged to to determine the stoichiometry of resident-protein loaded into ND from the total scattering intensity measured during SAXS. However, the current methods of determination of MM from SAXS make assumptions that are only valid for samples with homogenous electron density and globular or compact topologies (128-130). The physical properties of ND violate these assumptions due to the heterogenous electron density and scattering contrast arising from three distinct densities (protein, lipid acyl tails, and phospholipid head groups) and their decidedly non-globular disc shape known to exhibit fluid-like motions (125, 131).

In general, calculation of the molecular mass (MM) of a protein or particle directly from solution SAXS intensity curves is widely used as it allows for determination of the oligomeric state of the biomolecule, as well as monitoring of aggregation or sample degradation. One method for the determination of MM directly from SAXS data is enabled by the power-law relationship between the ratio between the term, Q_R, and particle mass as defined by Rambo and Tainer (129),

$$mass = \left(\frac{Q_R}{e^c}\right)^{1/k} \tag{5.3}$$

where, Q_R represents the ratio of the I(0) to the total scattering intensity divided by R_g, e is Euler's number (e = 2.718), and *k* and *c* are experimentally determined constants specific to a class of macromolecule (e.g. protein, RNA, etc.) calculated by analysis of large sets of experimental SAXS data. Unfortunately, the dependence upon the empirical constants *k* and *c*, limits MM determination via this method to protein only and RNA only particles (129), and therefore is not suitable for use in calculating the MM of a nanodiscs or resident-protein-ND complex due to a lack of experimental data from which to determine constants *k* and *c*.

Though considerably more challenging than the aforementioned method due to the need for accurate knowledge of the concentration of the particle within the solution and its partial specific volume ($\bar{v}$), the more classical method for determination of MM by SAXS is more widely applicable (128, 130). By this method, in general, MM of particles in solution is calculated by from the absolute scattering at the zero-angle, or I(0). Notably, I(0) cannot be directly measured as zero-angle scattering cannot be distinguished from the unscattered radiation of the direct beam, thus it must be extrapolated by the Guinier approximation (126) as delineated above. The extrapolated I(0) values must then been placed on an absolute scale by comparison to known MM protein standards (*i.e. l*ysozyme) or

by its relationship to the absolute scattering from water (128). Utilization of water as a secondary standard is accomplished by dividing the relative I(0) for the particle from the experimental constant scattering of water measured at the time of the particle SAXS experiment. Multiplication of this value by the absolute scattering of water to yields the absolute I(0) for the particle of interest. With knowledge of the absolute I(0) for the particle it is possible to determine of the MW, assuming accurate knowledge about the concentration particle in solution and its partial specific volume ($\bar{v}$) (130). Section 5.3.3 of this chapter details the development of an application of this method towards the determination of the MM of ND and protein-loaded ND ultimately to provide an independent measure of the number of target-proteins loaded into a ND.

5.2 Materials and Methods

5.2.1 ND Constructs used in SAXS and analytical ultracentrifugation (AUC) experiments

All constructs of empty ND, SUMO-tagged *C.concisus* PglC loaded ND , and ND with varying amounts of UndP added to the lipid component prepared and provided for SAXS and AUC experiments by Sonya Entova and Dr. Jean-Marie

Swiecicki (Imperiali Lab, MIT). All constructs used in the experiments presented within this chapter are summarized in table 5.1.

Table 5.1. Table of constructs and experimental concentrations for SAXS, AUC, and SEC-SAXS experiments.

ND Construct	Experiment	Conc., mg/mL
Empty ND		
MSP1D1 (10 nm)	SAXS	0.40
MSP1D1ΔH5 (8 nm)	SAXS	0.36
	AUC	0.20 - 0.60
	SEC-SAXS	*0.73 (0.25)
MSP1D1E3D1 (12 nm)	SAXS	0.40
Protein-loaded ND		
MSP1D1ΔH5 + SUMO-PglC	SAXS	0.36
UndP added to ND lipids		
MSP1D1ΔH5 + 1.7% UndP	AUC	0.62
MSP1D1ΔH5 + 3.3% UndP	AUC	0.51
MSP1D1ΔH5 + 5.0% UndP	AUC	0.51
MSP1D1ΔH5 + 6.7% UndP	AUC	0.62
MSP1D1ΔH5 + 8.3% UndP	AUC	0.55
MSP1D1ΔH5 + 10% UndP	AUC	0.60
	SAXS	0.68
	SEC-SAXS	*0.94 (0.32)

* Numbers in parentheses denote the effective concentration at the beam for SEC-SAXS samples which experience ~2/3 dilution on the SEC column.

To prevent sample deterioration and aggregation, all SAXS samples were kept cold, but never frozen, throughout the entire experimental process by shipping on wet ice to the beamline, maintenance of temperature by storage at 4

°C prior to experimental run, and experimental temperature control at 10 °C, unless otherwise noted.

5.2.2 Static Solution SAXS Data Collection and Processing

Solution SAXS experiments of ND samples were performed at beamline 4-2 at the Stanford Synchrotron Radiation Lightsource (SSRL) (Menlo Park, CA). ND samples were analyzed at the concentrations outlined in table 5.1 and with reserved gel-filtration matched buffer for each sample serving as the buffer blank for subtraction. To minimize aggregation-propensity and phase-transition within the ND lipids, samples were held at 283.15 K for experiments using a water-chilled Autosampler block and capillary holder sample environment at the beamline. For each data set, 30 μL of sample was was drawn into the experimental capillary via the Autosampler needle and oscillated during data collection to reduce radiation damage. Ten images with one second exposure at λ = 1.12 Å were collected on a Pilatus 3 x 1M detector positioned 0.45 m from the sample resulting in a resolution range of $0.0223 > q = 4\pi\sin\theta/\lambda > 1.69$ Å^{-1}, where 2θ is the scattering angle.

Data images were scaled and radially integrated automatically via SAXSPipe implementing the program sastool at the beamline. Scattering profiles computed for each image were subsequently averaged and outliers with $\chi 2 \geq 2.0$

were removed with SAXSPipe. The buffer contribution to the sample scattering was removed by subtraction of the scatting from the matched buffer sample form the sample to give the one-dimensional scattering intensity profile for the ND samples. Further processing and analysis of the SAXS data was carried out with the ATSAS suite of programs (132). Radius of gyration (R_g) for each sample was determined from transformation of the scattering intensity profiles in PRIMUS (133) via the Guinier approximation (126) (equation 5.1) within a fitting limit of $qR_g \leq 1.3$. The following fitting limits were used for each sample: MSP1D1, qR_g = 0.652 - 1.278 $Å^{-1}$; MSP1D1ΔH5, qR_g = 0.765 - 1.30 $Å^{-1}$; MSP1D1ΔH5 + PglC, qR_g = 0.49 - 1.29 $Å^{-1}$; and MSP1D1ΔH5 + 10% UndP, qR_g = 0.47 - 1.28 $Å^{-1.}$ The particle distiance distribution functions, P(r), for each sample were calculated from the scattering intensity curves in PRIMUS using the GNOM and autoGNOM tools to determine the longest inter-particle vector, D_{max}. For P(r) analyses the following q-ranges of data were used: empty 8nm ND = $0.0252 \leq q \leq 0.1796$ $Å^{-1}$; PglC loaded 8 nm ND = $0.0152 \leq q \leq 0.1293$ $Å^{-1}$; and empty 8nm ND with 10% UndP added = $0.0106 \leq q \leq 0.1614$ $Å^{-1}$.

5.2.3 Creation of molecular model of SUMO-SGSG-PglC in MSP1D1 (10 nm) ND

Theoretical molecular models of SUMO-PglC incorporated into ND were generated by first creating a predicted model for the SUMO-SGSG-PglC fusion construct using the ultra-deep learning structure prediction algorithm, RaptorX (49). The predicted model was then overlaid with the experimentally determined model for PglC described in Chapter 2 to determine SUMO-tag association with the experimentally determined structure for PglC. The SUMO-SGSG-PglC model of PglC was docked into a lipid bilayer slab by the PPM server (72) to determine the likely membrane interaction for the theoretical fusion construct. The calculated membrane plane was manually aligned with the membrane (chain C) of the molecular MSP1D1 DOPC nanodisc model (PDB ID: 2MSC, chains A and C) to position PglC into the ND model. For the model with 2 SUMO-PglC/ ND, the same manual alignment was repeated on the opposite face of the ND lipid bilayer. The relative size of the model of PglC requires that only one SUMO-PglC can be resident on each leaflet of the bilayer.

5.2.4 *Calculation of partial specific volumes for MSP, SUMO-PglC, and DMPC and DMPG lipids*

SEDNTERP(134) (http://sednterp.unh.edu/) was used to calculate the partial specific volumes and molecular masses for the MSP constructs of the ND, as well as for SUMO-PglC. The following sequences were input for calculation:

Table 5.2. AA sequences and output partial specific volumes ($\bar{v}$) from SEDNTERP.

Protein Construct	AA sequence	MW (kDa)	$\bar{v}$, 25 °C ($cm^3\ g^{-1}$)	$\bar{v}$, 10 °C* ($cm^3\ g^{-1}$)
MSP1D1Δ H5	GHHHHHHHDYDIPTTENLYFQGSTFSKLREQL GPVTQEFWDNLEKETEGLRQEMSKDLEEVKAK VQPYLDDFQKKWQEEMELYRQKVEPLGEEMRD RARAHVDALRTHLAPYSDELRQRLAARLEALK ENGGARLAEYHAKATEHLSTLSEKAKPALEDL RQGLLPVLESFKVSFLSALEEYTKKLNTQ	22.1	0.7257	0.7193
I64M/I94M SUMO-SGSG-PglC	MGHHHHHHGSLQDSEVNQEAKPEVKPEVKPET HINLKVSDGSSEIFFKIKKTTPLRRLMEAFAK RQGKEMDSLRFLYDGIRIQADQAPEDLDMEDN DIIEAHREQIGGSGSGMYRNFLKRVIDILGAL FLLILTSPIIIATAIFIYFKVSRDVIFTQARP GLNEKIFKMYKFKTMSDERDANGELLPDDQRL GKFGKLMRSLSLDELPQLFNVLKGDMSFIGPR PLLVEYLPIYETQKHRHDVRPGITGLAQVNGR NAISWEKKFEYDVYYAKNLSFMLDVKIALQTI EKVLKRSGVSKEGQATTEKFNGKN	35.8	0.7413	0.7349

* Temperature correction for $\bar{v}$ at the experimental temperature (T = 10 °C) given by equation 5.4 (135, 136).

Temperature correction for partial specific volumes of all components of the ND, including loaded protein, at the experimental temperature (T = 10 °C) was achieved with the following equation:

$$\bar{v}_T = \bar{v}_{25°C} + 4.25\times10^{-4}\,(T - 25) \quad (5.4)$$

The partial specific volumes of the DMPC and DMPG lipids were calculated from the literature values (137) of the specific molecular volumes, *V*, and molar masses, *M*, for DMPC (V_{DMPC} = 1.094 x 10^{-21} ml, M_{DMPC} = 659.0 g mol^{-1}) and DMPG (V_{DMPC} = 1.094 x 10^{-21} ml, M_{DMPC} = 666.9 g mol^{-1}) by the following equation (138):

$$V_{lipid} = \bar{v}_{lipid}\times\frac{M_{lipid}}{N_A} \quad (5.5)$$

Following $\bar{v}$ temperature correction by equation 5.4, $\bar{v}_{DMPC} = 0.9861\ cm^3\ g^{-1}$ and $\bar{v}_{DMPG} = 0.9699\ cm^3\ g^{-1}$ were obtained to allow for calculation of the partial specific volume of nanodiscs outlined in section 5.3.3 (equations 5.9, 5.10).

5.2.5 Sedimentation Velocity Analytical Ultracentrifugation (SV-AUC) data collection and processing

Sample preparation and data collection was performed by Sonya Entova and Dr. Jean-Marie Swiecicki (Imperali Lab, MIT). Interference and absorbance AUC data were processed via the SEDFIT/SEDPHAT program suite (139). The appropriate data range for the sample cell was fit to a Lamm equation model for a continuous C(s) distribution to determine molecular weight and relative

population size of particles in the sample. Partial specific volumes for the ND samples calculated according to equations 5.9 and 5.10 and buffer density of 1.007 calculated by SEDNTERP for the experimental buffer (134) and frictional ratio (*f/f0* = 1.3) were adjusted for during model fitting with non-linear regression. Goodness of fit was judged by a minimization of RMSD between the model and experimental data.

5.2.6 Size- Exclusion Chromatography-SAXS (SEC-SAXS)

SEC-SAXS experiments were performed at beamline 4-2 (BL4-2) at SSRL (Menlo Park, CA). As temperature control is not available for SEC-SAXS measurements at BL4-2, SEC-SAXS experiments were conducted at RT (~ 20 °C). The SEC column (GE Superdex 200 3.2/300, 2.4 mL column volume, 0.8 mL void volume) used in the experiments was equilibrated with 5 mL of degassed sample buffer prior to sample loading. ND samples (table 5.2) were centrifuged at 10.2 krpm for 30 minutes at 4 °C prior to loading onto the column. Sample injection was performed with 90 μL of each through the injection loop onto the SEC column. Injection and sample run were conducted at a flow rate of 0.05 mL /minute, resulting in a column pressure of 0.75-0.8 mPa. Images with one second exposure at λ = 1.12 Å were collected continuously through the run on a Pilatus 3 x 1M detector positioned 0.45 m from the sample in a detector distance resulting in a

resolution range of $0.0223 > q = 4\pi\sin\theta/\lambda > 1.69$ Å^{-1}, where 2θ is the scattering angle. The first 100 images collected were integrated, scaled, and averaged in sastools to create a representative matched buffer blank for the sample. R_g and I(0) plots for each each image were created by the fplcplots script and every five images over the run were averaged by the averagefplc script implemented at the beamline. To mitigate radiation-induced aggregation observed in the first sample run, the second sample beam shutter was intentionally left closed until the peak corresponding to the population of interest was observed on the FPLC (fast protein liquid chromatrography) UV (ultraviolet) absorbance trace. For this sample the buffer-blank was created from the last 100 frames after the UV absorbance signal returned to baseline. Analysis of the R_g and I(0) plots allowed for selection frames corresponding to the ND population of interest.

5.3 Results and Discussion

5.3.1 *Nanodiscs are highly-regular native-like lipid eviornments suitable for SAXS*

To confirm the homogeneity of the ND, and therefore their utility for biophysical characterization of PglC in a native-like lipid environment via solution SAXS, the scattering profiles for empty nanodiscs of two sizes 10 nm (MSP1D1) and 8 nm (MSP1D1ΔH5) discs were determined (figure 5.3A). It is important to

note that solution SAXS is a difference method defined by contrasts in the scattering of x-rays between different classes of molecules (e.g., protein vs. solvent). Due to the negative scattering length density of the hydrocarbon tails of the lipids relative to the aqueous solvent, the scattering from lipids is highly similar to the scattering of water, therefore the scattering contribution from the lipids is effectively lost during the buffer subtraction (125). This leads to the characteristic observation of an initial minimum at q = 0.07 $Å^{-1}$ feature observed in the scattering intensity of nanodiscs (figure 5.3A, arrow).

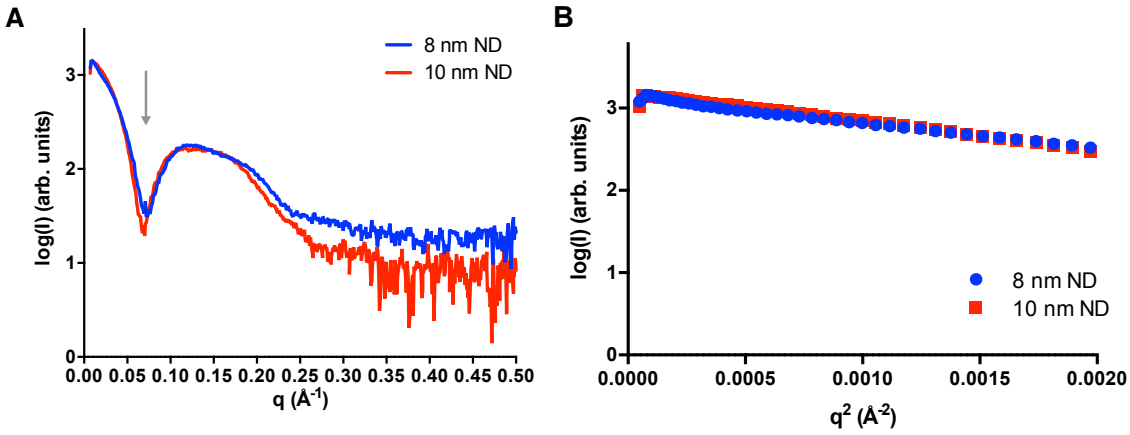

Figure 5.3. Solution SAXS analyses for 8 nm and 10 nm empty ND. A, Scattering intensity profiles for 8 nm and 10 nm empty nanodiscs. 8 nm ND scattering intensity depicted as blue line. 10 nm ND scattering intensity data depicted as red line. Characteristic initial minimum observed in solution SAXS intensity profiles denoted with gray arrow. **B,** Guinier analysis of 8 nm (blue circles) and 10 nm ND (red circles).

5.3.2 PglC incorporation into ND is observable by SAXS

Analysis of the SAXS data clearly illustrates ideal homogenous nature of the ND for use in the study of membrane-proteins in a more native-like lipid environment by solution SAXS. As such, similar experiments were repeated with 8 nm ND loaded with SUMO-SGSG-*C. concisus* PglC (SUMO-PglC) to confirm the observation of loaded protein within discs by SAXS. Gel densitometry analyses performed by Sonya Entova and Dr. Jean-Marie Swiecicki (Imperiali Lab, MIT) suggests a stoichiometry of 2 SUMO-PglC per 8 nm disc (data not shown). In the 8 nm discs (MW= 93 kDa including lipid) (140), SUMO-PglC (35.8 kDa) would account for 46% of the total scattering mass of the loaded ND particles. This is well above the theoretical limit of 20% detection (141). As a result, it was expected that sufficient signal required to resolve resident SUMO-PglC within the ND would be achieved. Analysis of the scattering intensity profile for the 8 nm ND with SUMO-PglC resident within the disc clearly shows a change in the inital minimum at q= 0.07 A^{-1} in comparison to the empty discs (figure 5.4A). The removal of this feature corresponds to the replacement of the effective negative scattering from the lipids within the disc with positive scattering from the SUMO-PglC resident protein.

Similarly, Guinier analysis of the SUMO-PglC loaded 8 nm ND samples showed sufficient linearity in the low-angle region for the determination of the R_g.

The radius of gyration for the SUMO-PglC loaded 8 nm ND (R_g = 57.4 ± 0.23 Å^{-1}) is significantly larger than the empty disc; this is expected from the increased scattering mass in the discs loaded with 2 SUMO-PGlC/ disc (figure 5.4B). Comparison of the pair-wise distance distributions for the empty and loaded ND shows a clear increase in in the maximum length the longest inter-particle of the vector (D_{max}) (figure 5.4C).

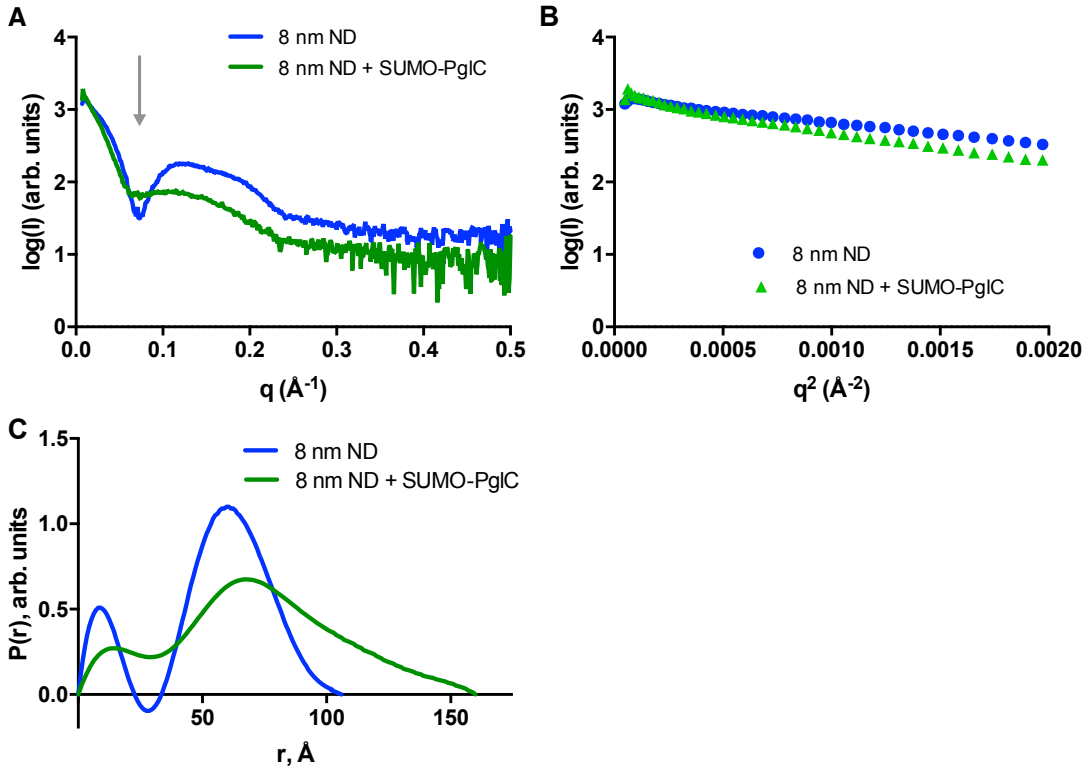

Figure 5.4. SAXS analysis of 8 nm ND empty and loaded with SUMO-PglC. A, scattering intensity profiles with the position of ND characteristic initial minima at ≅ 0.07 Å^{-1} indicated by gray arrow. **B,** Plot of Guinier region. **C,** Plot of P(r) analyses.

Initially, the increased D_{max} vector was accounted for by hydrophobic protein-protein interactions between the amphipathic helix identified during initial structure predictions for PglC (17) and the inner surface of the MSP protein belts encompassing the ND lipid bilayer. However, the now experimentally determined structure of PglC and the availability of an atomistic model for a similar ND (PDB ID: 2MSC) (142) enabled an estimation of the longest inter-particle vector from the molecular coordinates (figure 5.5). If 2 PglC are resident in each ND as suggested by gel densitometry, the longest protein-protein interparticle vector estimated from the model coordinates corroborates the experimentally determined increase vector length for a loaded ND in comparison to an empty ND (figure 5.5C).

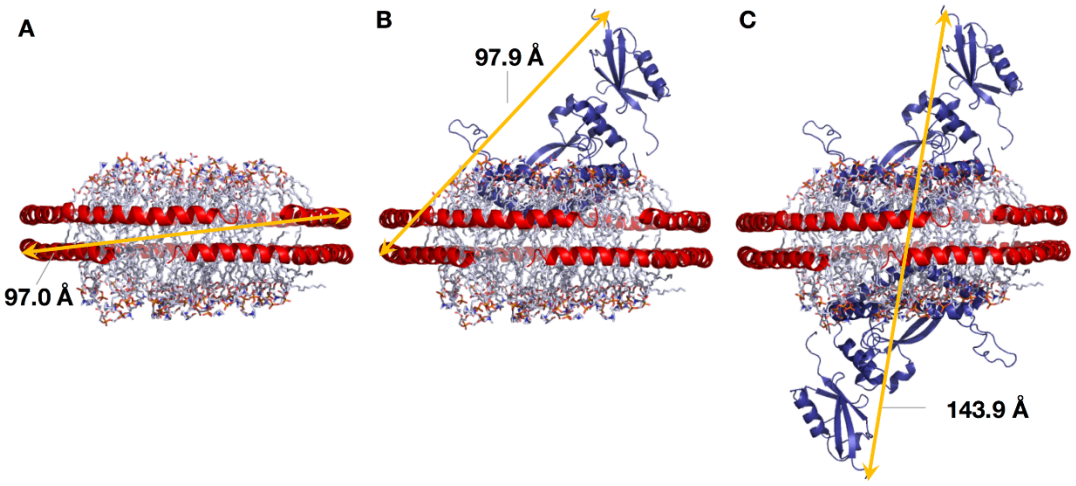

Figure 5.5. Estimation of D_{max} vector from the theoretical model of PglC in ND. **A,** The longest protein-protein inter-particle vector in the empty ND model (PDB ID 2MSC, chains A and C) measures to be 97.0 Å. **B,** The longest protein-protein inter-particle vector in the ND model with one PglC per disc measures to be 97.9 Å. **C,** The longest protein-protein interparticle vector in the ND model with two PglC per disc measures to be 143.9 Å.

5.3.3 Quantification of resident-protein to ND stoichiometry from SAXS data

Thus based on this model, it is critical that the stoichiometry of SUMO-PglC within the ND is two SUMO-PglC per ND to explain the increase in the D_{max} vector observed experimentally. Typically, one common method to determine stoichiometry of protein-loading in the ND is thorough utilization of gel densitometry by quantitative Coomasie staining (21). However, due to variability in binding of Coomasie to proteins gel densitometry does not necessarily afford a

high precision of protein quantification (143). Therefore, in effort to provide an independent measure for resident-protein to disc stoichiometry, the subtractive power of SAXS can be leveraged to directly calculate the number loaded proteins within the disc from the scattering data. To this end, the following mathematical method was developed, where the total scattering mass of protein loaded in the ND is calculated and then divided by the MW for the resident protein to obtain relative quantitation of the resident protein in the ND.

To calculate the molecular mass of PglC resident in the MSPΔH5 nanodiscs via the total scattering mass the molecular mass of the empty ND was subtracted from the molecular mass of the loaded nanodiscs calculated according to the equation described by Mylonas and Svergun (130),

$$MM = \frac{\left[N_A\left[\frac{I(0)}{c}\right]\right]}{\Delta\rho_M{}^2} \tag{5.6}$$

where, N_A is Avogadro's number (6.023 x 10^{23} mol^{-1}), I(0)/c is the zero-angle scattering intensity normalized by sample concentration, and $\Delta\varrho_M$ is the scattering contrast per mass. $\Delta\varrho_M$ defined by Mylonas and Svergun (130) accounts for by the scattering density from the macromolecule and the solvent,

$$\Delta\rho_M = \left[\rho_{M,prot} - (\rho_{solv}\bar{v})\right]r_o \tag{5.7}$$

where, $\varrho_{M,prot}$ is the number of electrons per mass of dry protein (3.22 x 10^{23} e g^{-1}), ϱ_{solv} is the number of electrons per volume of the aqueous solvent (3.34 x 10^{23} e cm^{-3}), $\bar{v}$ is the partial specific volume of the protein, and r_o is the classical scattering length of one electron (2.8179 x 10^{-13} cm). The heterogeneous composition of nanodiscs, both protein-loaded and empty, must be accounted for in the calculation of the partial specific volume, $\bar{v}$. The partial specific volume of empty nanodiscs, $\bar{v}_{ND}$, was calculated as defined by Inagaki, Ghirlando, and Grisshammer (144),

$$\bar{v}_{ND} = \frac{2M_{MSP1D1\Delta H5}\bar{v}_{MSP1D1\Delta H5} + xM_{DMPC}\bar{v}_{DMPC} + yM_{DMPG}\bar{v}_{DMPG}}{2M_{MSP1D1\Delta H5} + xM_{DMPC} + yM_{DMPC}} \quad (5.8)$$

where, M represents the molar masses of the nanodisc components, $\bar{v}$ represents the partial specific volumes of the components. The variables x and y represent the relative stiochiometries of the lipids DMPC and DMPG in the nanodiscs, respectively. To calculate the partial specific volume for protein-loaded nanodiscs, $\bar{v}_{ND,prot}$, the above equation can be modified by adding mass (M_{PglC}), partial specific volume ($\bar{v}_{PglC}$) , and expected stoichiometry terms (z) for the loaded protein to give

$$\bar{v}_{ND,prot} = \frac{2M_{MSP1D1\Delta H5}\bar{v}_{MSP1D1\Delta H5} + xM_{DMPC}\bar{v}_{DMPC} + yM_{DMPG}\bar{v}_{DMPG} + z\,M_{PglC}\,\bar{v}_{PglC}}{2M_{MSP1D1\Delta H5} + xM_{DMPC} + yM_{DMPC} + z\,M_{PglC}} \quad (5.9)$$

Thus, to calculate the $\bar{v}_{ND,PglC}$ and $\bar{v}_{ND}$ for the 8 nm ND used in these experiments, the number of lipids loaded in the 8 nm disc was assumed to be 100 with a 3:1 DMPC:DMPG ratio, therefore, $x = 83$ and $y = 27$ (145). Using this stoichiometry and the M_{lipids} and $\bar{v}_{lipids}$ *calculated with SEDNTERP*, $\bar{v}_{ND,PglC} = 0.7343$ cm^3 g^{-1} and $\bar{v}_{ND} = 0.7339$ cm^3 g^{-1}.

Additionally, to accurately compute the MM, the forward scattering of the nanodiscs, Abs$I(0)_P$, at zero-angle must be calculated in the absolute scale through the use of water as a secondary standard (128) by,

$$AbsI(0)_P = \frac{I(0)_{P,exp}}{I(0)_{w,exp} \times \frac{d\Sigma}{d\Omega_{q=0}}} \tag{5.10}$$

where, $I(0)_{P,exp}$ is the experimental forward scattering of the nanodiscs extrapolated from the Guinier approximation of the experimental data, $I(0)_{w,exp}$ is the experimental forward scattering of water calculated through subtraction of scattering from the empty flow-cell and flow-cell filled with water (130), and $d\Sigma/d\Omega_{q=0}$ is the absolute forward scattering intensity of water at the experimental temperature as defined by Orthaber, Bergmann, and Glatter (128),

$$\frac{d\Sigma}{d\Omega_{q=0}} = \rho^2 k_b T \mathrm{X}_T \tag{5.11}$$

where, ρ is the scattering length density of water (9.388 x 10^{10} cm^{-2}) , k_b is the Boltzmann constant (1.3806 x 10^{23} m^2 kg s^{-2} K^{-1}), and X_T is the isothermal

compressibility of water at experimental temperature, 283.15 K, (4.78×10^{-10} Pa^{-1}) as extrapolated from the isothermal compressibility of water at 288 K determined by Kell (146).

Using the aforementioned approach with experimental solution SAXS data collected for both empty and loaded 8 nm ND discussed in section 5.3.2 and biophysical parameters described in table xxxx resulted in the calculation of a stoichiometry of ~ 2 PglC/ND as outlined in table 5.4. The slight overestimation of the stoichiometry (~2.4) can be explained by the uknown mass of the lipids displaced by the protein loaded into the disc. Ultimately, this novel method for determining the stoichiometry of resident protein in a ND via the total scattering mass independently confirms the ~2 SUMO-PglC/ ND as observed by gel densitometry. As the collection of a water blank dataset is common practice for all SAXS experiments, the mathematical method described above enables the independent calculation of resident protein stoichiometry for any high-quality membrane-protein loaded-ND solution SAXS data collected.

Table 5.3. Table of parameters calculated to determine SUMO-PglC - ND stoichiometry.

Parameter		Value
Partial specific volumes ($cm^3\ g^{-1}$)		
	DMPC, $\bar{v}_{DMPC}$	0.9699
	DMPG, $\bar{v}_{DMPG}$	0.9861
	8 nm ND MSP, $\bar{v}_{MSP}$	0.7193 *
	PglC, $\bar{v}_{PglC}$	0.7345 *
	Empty 8 nm ND, $\bar{v}_{ND}$	0.7340
	SUMO-PglC loaded 8 nm ND, $\bar{v}_{ND,PglC}$	0.7344
Constants		
	Avogadro's nubmer, N_A (mol^{-1})	6.02×10^{23}
	No. electrons/ mass dry protein, ρM_{prot} ($e\ g^{-1}$)	3.22×10^{23} **
	No. electrons/ vol. aq. solvent, ρM_{solv} ($e\ cm^{-3}$)	3.34×10^{23} **
	Scattering length of an electron, r_o (cm)	2.82×10^{-13} **
Scattering from water		
	Experimental constant scattering of H_2O, $I(0)_{H2O,exp.}$ (cm^{-1})	5.42×10^{2}
	Abs. scattering of H_2O at 283.15 K (cm^{-1})	1.64×10^{2} **

* calculated from aa sequence by SEDNTERP; ** value reported by Orthaber, Bergmann, and Glatter (128).

Table 5.4. Experimental and computed parameters for the empty 8 nm ND and SUMO-PglC loaded ND.

Sample	Conc. (10^{-3} g cm^{-3})	I(0) (cm^{-1})	I(0)/c (10^{6} cm^{2} g^{-1})	AbsI(0) $_{283.15\ K}$ (10^{-2} cm^{-1})	AbsI(0)/c $_{283.15\ K}$ (cm^{2} g^{-1})	$\Delta\rho_M$ (10^{10} cm g^{-1})	$MM_{283.15\ K}$ (kDa)
Empty ND							
8 nm ND, [h]	0.358	624	1.74	0.19	52.74	0.802	494.0
8 nm ND, [l]	0.178	288	1.62	0.01	48.96	0.802	458.6
						MW average =	**464.7**
Loaded ND							
8 nm ND + SUMO-PglC, [h]	0.18	723	4.02	0.22	121.54	1.12	587.2
8 nm ND + SUMO-PglC, [l]	0.09	334	3.71	0.10	112.29	1.12	542.6
						MW average =	**564.9**
ΔMW*							86.42
SUMO-PglC**							35.8
						SUMO-PglC/ ND	**2.4**

Notation: [h] indicates high concentration samples, [l] indicates low concentration samples; *$\Delta MW = MW_{avg,empty} - MW_{avg,loaded}$; ** MW of SUMO-PglC calculated from the AA sequence by SEDNTERP (see methods section).

5.3.4 Inclusion of UndP in ND results in sample heterogeneity

Towards understanding of the effect of PglC membrane association in the presence of its native membrane-bound substrate, UndP, empty 8 nm ND were prepared with the addition of 10% UndP. Inclusion of 10% UndP into the lipid mixture for ND preparation results in incorporation of approximately 3 UndP per ND. Notably, scattering intensity profiles curves remain similar for discs with and without the addition of UndP as expected; however, calculation of the P(r) function revealed an unexpected, drastically increased D_{max} for the ND particles with 10% UndP incorporation (figure 5.6C). Analysis of the Guinier plot revealed linearity in the low-angle region consistent with high-quality non-aggregated samples (figure 5.6B); thus, aggregation can be excluded as a cause for the unexpected increase in D_{max} observed (127) . As a result, despite appearing homogeneous by dynamic light scattering (DLS) measurements obtained by Sonya Entova and Dr. Jean-Marie Swiecicki (Imperiali Lab, MIT) (data not shown), heterogeneity within the sample in the form of multiple distinct populations of particles (e.g. ND and lipid micelles) versus non-specific aggregation was hypothesized to give rise to the increased D_{max} measurement.

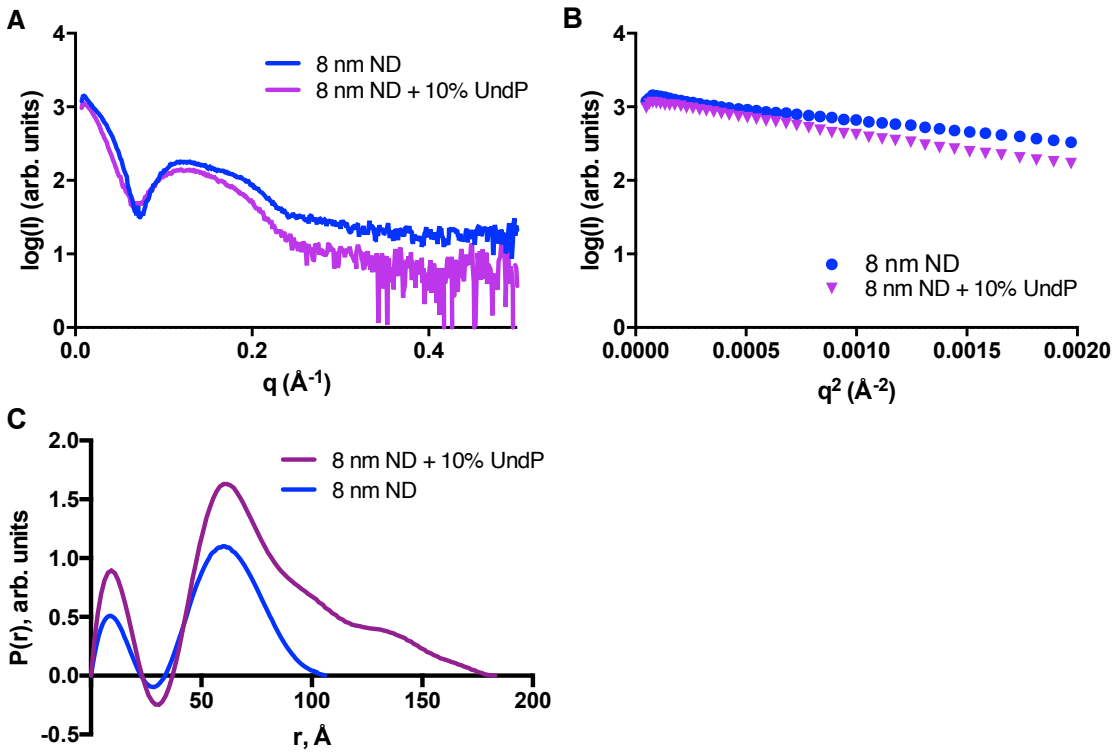

Figure 5.6. Analysis of scattering data for 8 nm ND ± 10% UndP added to the lipid component. A, Scattering intensity profiles with 8 nm ND data shown as blue line and 8 nm ND with 10% UndP added shown as purple line. **B,** Scattering data displayed in a Guinier plot for 8 nm ND shown as blue circles and scattering data for 8nm ND with 10% UndP added shown in purple triangles. **C,** P(r) curve for 8 nm ND shown as blue line and for 8 nm ND with 10% UndP added shown as purple line.

To confirm the heterogeneity suspected within the ND samples with added UndP, the samples were analyzed by analytical ultracentrifugation (AUC). AUC, in particular sedimentation velocity (SV)-AUC, offers increased resolution of hydrodynamic radii over DLS (144), therefore allowing for the deconvolution of particle populations outside the resolution limit of standard DLS experiments

(147). Fitting of the collected interference SV-AUC data with to a Lamm Equation model for continuous C(s) distributions (144) established that the empty nanodiscs behave as a single population with a sedimentation coefficient (s) of ~2.5 representing 82% of the mass in the sample (figure 5.7A,B). Oppositely, fitting of the data collected from ND samples with 10% UndP added revealed clear heterogeneity in the sample with the largest population (s ≅ 2.5) only accounting for 45-47% of the total mass (Figure 5.7B). The increased heterogeneity appears to be dependent upon the amount of UndP added with discs with < 5% UndP appeared to behave similar to the empty discs with no added UndP (figure 5.7C). At concentrations of UndP <5% incorporation of UndP into the discs is sub-stoichiometric, therefore the major population sedimenting at s ~2.5 is representative of empty discs with no incorporated UndP.

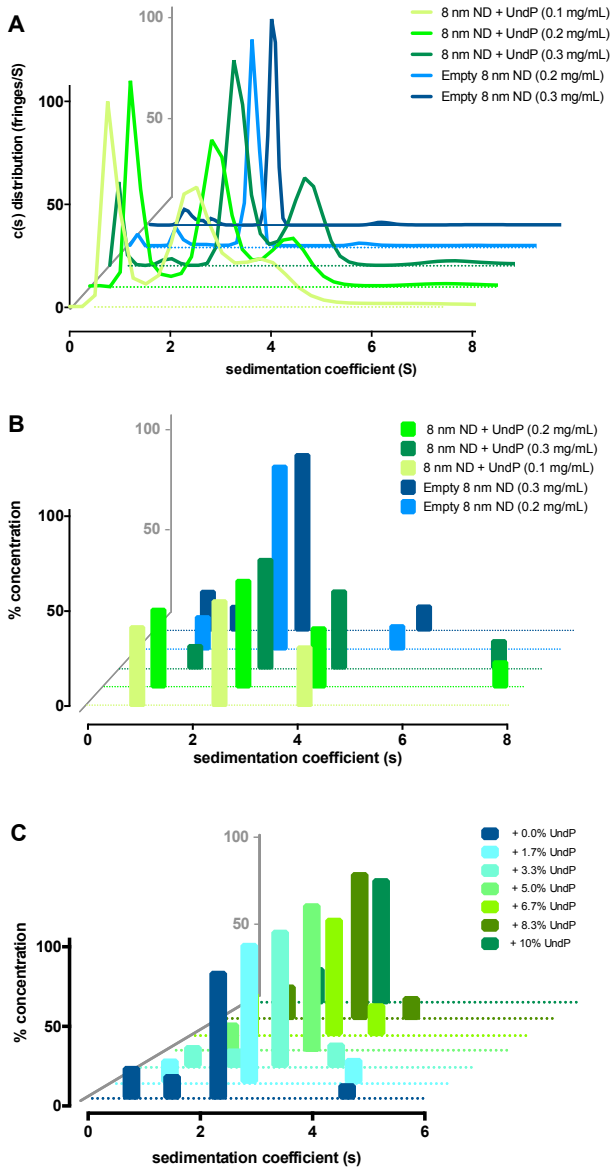

Figure 5.7. C(s) distributions modeled from interference SV-AUC data for ND samples with varying amounts of UndP added to the lipid component. A, Multiple populations are seen with the addition of 10% UndP in a concentration independent manner. **B,** Percent concentration of the total mass within the sample of each C(s) peak from panel A. **C,** Effect of added UndP concentration illustrated by plots percent concentration of the total mass within the sample of each C(s) peak for samples with 0.0% to 10% UndP added. All plots created in GraphPad Prism.

5.3.5 SEC-SAXS as a method for characterizing a single ND population

The SV AUC experiments illustrate the need for an additional purification step to isolate a single population for biophysical experiments involving the physiologically-relevant concentration of 10% UndP added per disc. Specifically, for SAXS experiments, the development of in-line size-exclusion chromatography (SEC) apparatuses enables collection of SAXS data from isolated populations directly as they come off the column. Preliminary SEC-SAXS experiments with the 8nm ND and 8 nm ND with 10% UndP allowed for sufficient separation of single populations (figure 5.8A). Continuous data acquisition during SEC-SAXS experiments allows for selection of frames in which only a single population is present based on the chromatogram for further data analysis. Notably, during these experiments a climbing I(qmin) trace (figure 5.8 B, blue circles) for the empty ND sample was observed. This behavior is indicative of radiation-induced aggregation coating the experimental capillary in the beam path. Therefore, the R_g measured for the empty 8nm ND by SEC-SAXS was significantly larger (R_g = ~53 Å) in comparison to that measured by static SAXS (R_g = ~45 Å) is likely resultant of over estimation of the R_g due to the aggregation observed (127) . To mitigate these effects in further experiments, the shutter was intentionally left closed until the peak corresponding to the major population of ND was observed

by UV absorbance at λ = 280 nm (figure 5.8A, arrow). As a result, the I(qmin) trace is not visible for the 10% UndP added sample. As a result, the preliminary SEC-SAXS data collected for the ND sample with 10% UndP added was of significantly higher quality. Guinier anlyais shows the R_g reported for this sample (R_g = ~ 56 Å) was highly similar to that measured statically, Rg = ~57 Å (figure 5.6B). Thus, sample heterogeneity cannot be the only cause for the apparent unexpectedly large D_{max} and R_g for the 8nm samples with approximately three UndP incorporated per ND. One such possibility is that the inclusion of UndP into the lipid-bilayer causes slight distortion and disassembly for the nanodiscs as was previously observed for the addition of cholate to ND (148).

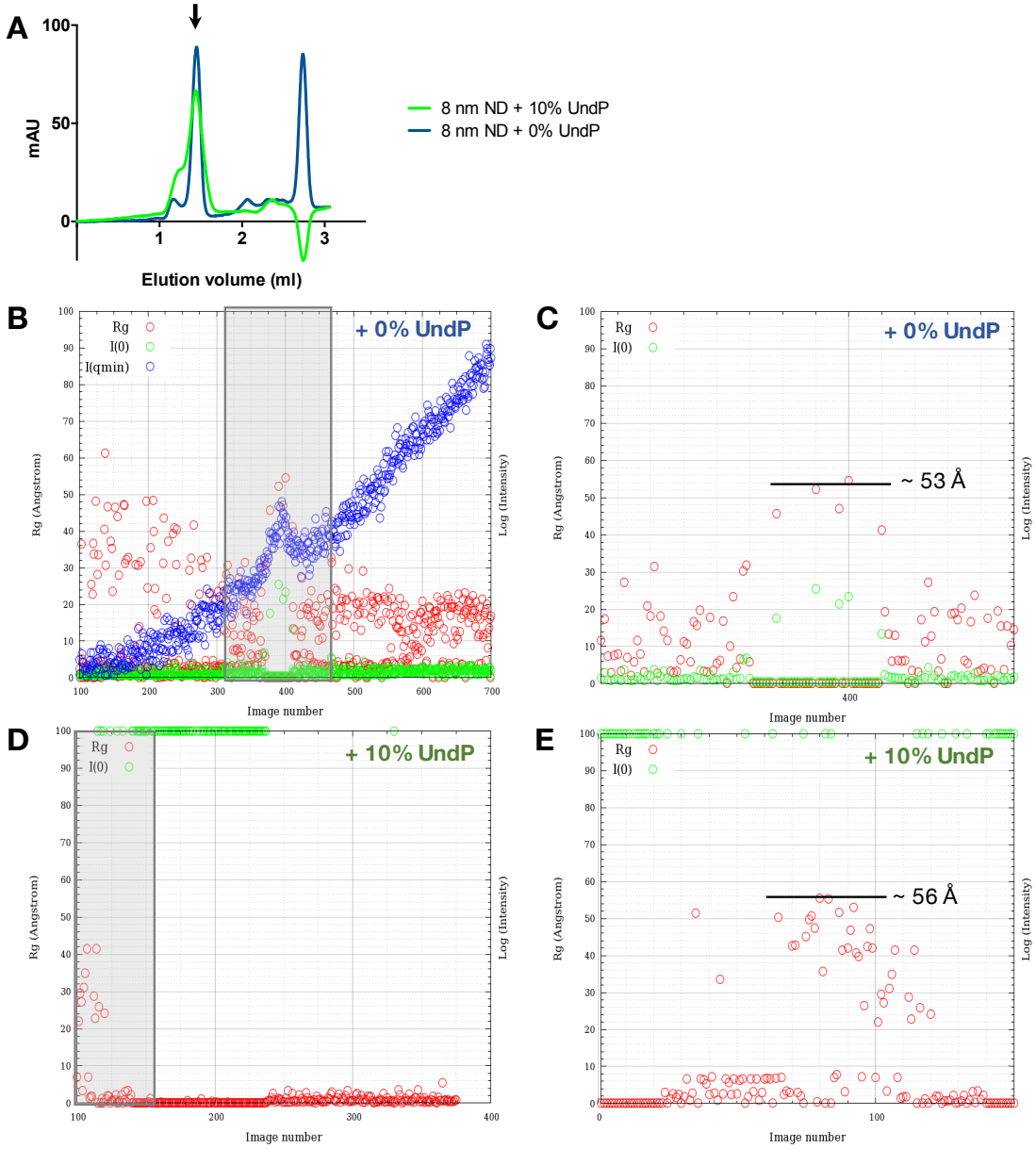

Figure 5.8. Preliminary SEC-SAXS experiments show separation between multiple ND populations. A, Absorbance (λ = 280 nm) FPLC trace from experimental runs of 8 nm ND with 10% UndP added to the lipid component of the discs (green trace) and empty 8 nm ND with no added UndP (blue trace). Arrow indicates peak of major population of interest. **B,** R_g and I(0) script for every image collected during SEC-SAXS run for ND with no UndP added. Plot created with beamline fplcplots. Red open circles indicate calculated R_g, green open circles indicate extrapolated I(0), blue open circles indicate measured intensity at the

smallest angles, I(qmin), and gray box indicates zoom region illustrated in panel C. **C,** R_g and I(0) for every frame within zoom region indicated in B. Black line indicates average R_g of the peak corresponding to the ND population of interest. **D,** R_g and I(0) script for every image collected during SEC-SAXS run for ND with 10% UndP added. Plot created with beamline fplcplots. Red open circles indicate calculated R_g, green open circles indicate extrapolated I(0), and gray box indicates zoom region illustrated in panel C. E, R_g and I(0) for every frame within zoom region indicated in B. Black line indicates average R_g of the peak corresponding to the ND population of interest.

5.4 Conclusions and Future Directions

From the data presented in this chapter, it is evident we are able to produce and characterize high-quality lipid ND for biophysical and structural studies of PglC in a native-like phospholipid environment. SAXS data for SUMO-PglC loaded into the disc clearly illustrates resolution of the 35.8 kDa particle within an 8 nm ND from the changes in the scattering intensity profile. Notably, incorporation of PglC into nanodiscs yielded an in initially unexpected increased D_{max} of the particles. However, upon modelling the expected SUMO-PglC loaded ND with 2 PglC/ND, the increased length was corroborated. With the atomic structure of PglC now known, it would be worthwhile, in the future, to attempt model-based shape reconstruction of SUMO-PglC in the ND via a similar hybrid model-based *ab initio* shape reconstruction method used to elucidate the structure of CYP3A4 embedded in a ND membrane (149).

Additionally, the mathematical method outlined in section 5.3.3 allows for the calculation of resident-protein loading into a ND from directly from high-quality SAXS data, thus, providing an independent corroboration of the loading of 2 PglC-SUMO per 8 nm ND as suggested by gel densitometry experiments. Notably, this mathematical method has only been tested with the SUMO-PglC loaded discs present in this chapter; thus, it would be worthwhile to confirm utility of the calculation on different ND samples like MSPE3D1 ND which preferentially have a 1:1 PglC:ND stoichiometry (21). However, preliminary SAXS experiments with these discs show high aggregation propensity, and therefore, optimization of experimental conditions will be required. Recent studies of terbium-mediated luminescence of PglC performed by Dr. Debasis Das (Imperiali Lab, MIT) (data not shown), suggests an additional independent quantification of PglC loaded into the ND could be achieved through quantification of luminescence upon saturating conditions of lanthanide as was observed for lanthanide-binding-tag (LBT)-Ubiquitin (150).

Lastly, towards understanding effects of substrate, Und, incorporation into ND with PglC, preliminary SAXS studies with empty discs with UndP added were performed. Though heterogeneity, confirmed by AUC, within the sample was originally hypothesized to be causative of an unexpected increase in D_{max} and R_g

for ND with 10% UndP added, preliminary SEC-SAXS experiments to isolate a single population of ND determined highly similar R_gs for the heterogeneous sample measured by static SAXS. Thus, the observed heterogeneity in the ND samples with 10% UndP added is not directly causitive of the increased molecular dimensions measured by SAXS. Toward understanding the increase particle dimensions of UndP addition, SEC-SAXS experiments on ND with varying concentrations of UndP added will be required to understand the effect of UndP addition into the ND. Notably, a similar phenomenon was observed for ND with incorporation of cholate where P(r) anlalyes from ND with increasing cholate concentration coupled with course-grained molecular dynamics simulations elucidated specific ND structural disruptions upon cholate incorporation into the ND (148).

CHAPTER SIX

Conclusions

In prokaryotes, protein glycosylation can be a determinant of pathogenicity as it plays a role in host adherence, invasion, and colonization (2, 3). Impairment of glycosylation in pathogenic bacteria leads to decreased virulence; thus, opening new avenues for the development of inhibitory agents (6). Notably, the first membrane-committed step of *en bloc* glycosylation pathways, marked by catalysis of the transfer of a phosphor-sugar from a soluble nucleotide-diphosphate sugar to a membrane-bound PrenP by a PGT enzyme, is a shared feature in the biosynthesis of many complex glycoconjugates (4).

PGTs are classified into two isofuncitonal, yet structurally- and mechanistically-diverse superfamilies. The polytopic PGT superfamily, typified by the structurally-characterized enzyme MraY, proceed through a ternary complex enzymatic mechanism and are known to be inhibited by the bi-substrate analog inhibitor tunicamycin (24). Conversely, the monotopic PGT superfamily is not inhibited by Tunicamycin although it catalyzes the same general reaction as the polytopic superfamily (20). Recently, Das *et al.* demonstrated PglC, an exemplar monotopic PGT, proceeds through a covalent-phospho-sugar

intermediate enzymatic mechanism (25). Prior to the work presented in this dissertation, the structure, and therefore the basis for lack of inhibition by tunicamycin, of the putative transmembrane protein monotopic superfamily was experimentally unknown. PglC, from the well-characterized N-linked glycosylation pathway in *C. jejuni*, shares significant homology to the PGTs, WbaP and PglB (*N.g.*) which catalyze the same membrane-committed steop in the O-antigen and O-linked glycoprotein biosynthetic pathways. The small size of PglC (201 aa) compared to homologous PGTs suggests it may represent the minimal catalytic unit for the monotopic PGT superfamily, thus positioning PglC as an optimal target for molecular structure elucidation (17). Notably, structure determination of PglC, poised to provide much-need structural information for an entire superfamily of PGT enzymes, will also add another example of the structure monotopic membrane protein to the vastly underrepresented class in the PDB.

Herein, the structure of *C. concisus* PglC including its putative TM domain has been solved to 2.74 Å resolution to reveal a novel protein fold with a unique α-helix-associated β-hairpin (AHABh) motif and largely solvent-exposed structure. Analysis of the structure reveals that both membrane-association and active site geometry are underpinned within the minimal structure of PglC by structurally-efficient fold-motifs in the form of proline kinks, short hydrogen

bonds, and extensive hydrogen-bond networks.

The RMH, confirmed through biochemical and bioinformatics studies, is a structurally defining feature of PglC and the monotopic PGT superfamily in that the hydrophobic helix is re-entrant on the cytoplasmic face of the membrane rather than membrane spanning as originally predicted. Comparison of the structural determinants membrane association evident for PglC with other monotopic membrane proteins suggests PglC represents a first-in-class structure of a novel membrane-interaction mode. The AHABh-motif represents a second structurally-defining feature of PglC. This motif is structurally and functionally conserved as it establishes the co-facial positioning of the catalytic Asp-Glu dyad. Molecular docking of PglC substrates, undecaprenyl phosphate (UndP) and UDP-N,N-diacetylbacillosamine (UDP-diNAcBac), within the active-site reveals co-incident binding sites, consistent with the proposed covalent phospho-sugar intermediate mechanism.

Lastly, loading of PglC into membrane-bilayer nanodiscs (ND) allows for the investigation of PglC structure and function within a more native-like membrane environment by small-angle x-ray scattering (SAXS). Detection of PglC laded in ND via SAXS confirms the application of the method for studying small, integral, monotopic membrane proteins in a membrane environment. In order to

independently corroborate biochemical estimations of the stoichiometry of PglC incorporation within ND, a mathematical approach by which resident-protein-to-ND stoichiometry can be deduced from measured scattering intensity was developed.

Overall, the understanding of the molecular structure and mode of membrane association of PglC presented in this dissertation and summarized in figure 6.1, provides critical understanding of the structure-function relationship within this pathway poised for inhibitor development.

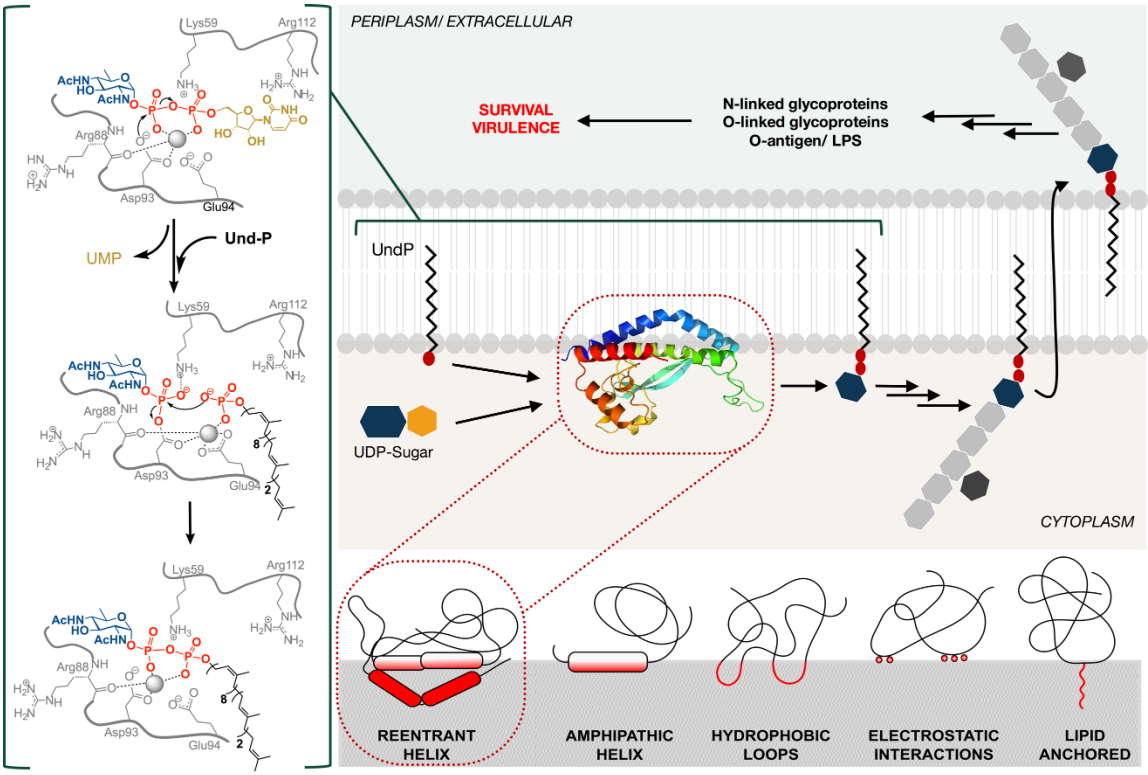

Figure 6.1. Structural basis for membrane-association and mechanism in the monotopic PGT superfamily, as exemplified by *C. jejuni* PglC. Green brackets denote the enzymatic mechanism for PglC supported by structural analyses. Red dashed boxes indicate the novel membrane-association mode observed for PglC.

APPENDIX A1

Understanding the feasibility of anomalous SAXS experiments

A1.1 Introduction

SAXS is a complementary technique in structural studies in that it provides unique structural information about a sample in solution, thus, it can offer lower-resolution context to a high-resolution structure of a protein constrained in the crystal lattice. SAXS structural studies on proteins with genetically-encoded lanthanide-binding tags (LBTs) (150-152) enables the measurement of anomalous small-angle x-ray scattering (ASAXS) intensity data by exploiting the high anomalous scattering of the lanthanides. ASAXS allows for the positions of the anomalous scatterer bound to the LBTs, often terbium (Tb) (150), to be located in relation to the protein, thus, providing distance information that would be highly sensitive to changes in conformation or relative orientations. For PglC, ASAXS experiments with PglC in nanodiscs could enable an understanding of the interactions between PglC and the next-downstream member in the N-linked glycosylation pathway, PglA, as well as a validation and further understanding of the relative orientation of PglC in the membrane.

In solution, anomalous scattering has two components cross-terms and anomalous terms. Cross-terms depend on the on the relative location of the anomalous centers in relation to the center of mass for the protein, while the anomalous terms depend on the the distances between the anomalous centers (153, 154). Using the iron anomalous signal from hemoglobin the cross-terms were observable to give iron to center-of-mass (CoM) of protein distance of ~8.6 Å, which is whin ~1 Å of the distance observed in a crystal structure (155). Notably, the second component terms, the anomalous terms, are at least two-orders of magnitude smaller than the cross-terms, consequently making their measurement likely beyond the the current experimental sensitivities.

An ASAXS experiment is highly similar to a SAXS experiment, but utilizes several wavelengths near the absorption edge of the anomalously scattering atoms instead of a fixed wavelength as in a standard experiment. The anomalous scattering is observable as scattering intensity differences at the various wavelengths. Precedent for ASAXS experiments have demonstrated successful measurement of the cross-terms between the heme iron and mygoblin from an anomalous scattering intensity difference of 1% of native myogoblin signal (155). Similarly, Tb cross-terms were measured in parvalbumin from anomalous scattering intensity differences of 5-7% (153). However, exploiting the high

anomalous scattering power of lanthanides (~ 20 scattering electrons) bound to genetically-encoded LBT tags should enable increased magnitudes of anomalous scattering intensity differences.

The work presented in this appendix details preliminary ASAXS experiments with a structurally well-characterized construct of ubiquitin with two genetically-encoded LBTs (156), and subsequent computational simulations of LBT-enabled ASAXS experiments to understand the feasibility of measurement of anomalous scattering intensity differences for both simple proteins and more complex systems.

A1.2 Materials and Methods

A1.2.1 Protein expression, purification, and $TbCl_3$-loading

Expression and purification of dLBT-Ubiquitin was performed following previously published protocols (152, 156). One liter Luria Broth (LB) media was inoculated with BL21(DE3) *E. coli* cells and allowed to grow at 37°C until an optical density of OD_{600} = 0.6-0.8. Protein expression was induced with 500 mg isopropyl-β-D-1-galacotpyranoside (IPTG) per liter of cultured cells. Induction was carried out at 16°C and cells were allowed to continue to grow until OD_{600} = 1.2-1.4. Cells were pelleted via centrifugation at 6000 rpm. The cell pellet was homogenized by

resuspension in Buffer 1 (50 mM Tris pH 8, 300 mM NaCl, 10 mM imidazole) at a volume of 5 mL/g of cell pellet. Cells were lysed using a microfluidizer. Cell debris was separated from the soluble fraction via ultracentrifugation at 38,000 rpm. dLBT-Ubiquitin was purified from the soluble fraction via immobilized nickel affinity column with elution via buffer 2 (100 mM NaCl, 20 mM HEPES pH 7, and 150 mM imidazole). Pure dLBT-Ubiquitin was dialyzed overnight at 4℃ against 4L of dialysis buffer (20 mM HEPES pH 7.5, 100 mM NaCl). Any aggregation was pelleted via ultra-centrifugation at 38,000 rpm and the resulting pure, soluble protein was diluted to 1 mg/mL in dialysis buffer. Purified dLBT-ubiquitin was separated into two aliquots. One aliquot was loaded with Tb^{3+} at a ratio of 1:1.1 (protein: $TbCl_3$) molar equivalents with two-minute incubations between each addition of $TbCl_3$. The reserved aliquot was loaded with Ca^{2+} via incubation with 1:1 protein to $CaCl_2$ for 1 hour on ice.

A1.2.2 Preparation of SAXS samples

Resulting pure protein was concentrated using an Amicon centrifugal concentrator with a molecular weight cut-off of 3,500 Da to a final concentration of 15.5 mg/mL. To ensure accurate buffer subtraction, concentrated dLBT-ubiquitin was dialyzed overnight at 4℃ against 4 L of dialysis buffer. An aliquot

of the resulting dialysis buffer was reserved for background SAXS measurements. Following dialysis, the dLBT-ubiquitin protein was flash frozen in liquid nitrogen for transport to the beamline. Protein was thawed on ice at the beamline and serial dilutions of Tb^{3-} loaded dLBT-ubiquitin were performed from 10 mg/mL to 0.88 mg/mL. ASAXS data collection was carried out with a dLBT-ubiquitin concentration of 5 mg/mL.

A1.2.3 ASAXS data collection and processing

Anomalous SAXS data collection was performed at Brookhaven National Lab (BNL), National Synchrotron Light Source-II (NSLS-II), beamline 16-ID-B (LiX), Upton, NY. 1 mM $TbCl_3$ in 25 mM ammonium acetate was used as a Tb standard to perform an x-ray absorption spectrum for Tb. the anomalous scattering factors for the absorption spectra were determined by CHOOCH (157). For anomalous SAXS measurements, data was collected at five different energies (7521 eV, 7517 eV, 7510 eV, 7500 eV, and 7482 eV) spanning five different Bragg angles, respectively (15.0885 ° , 15.0965 ° , 15.111 ° , 15.132 ° , and 15.169 °). The SAXS scans at each energy were averaged and buffer subtracted at the beamline with PyXS (https://www.bnl.gov/ps/x9/software/pyXS.asp). Resulting buffer-subtracted data were normalized to water peak scattering intensity, q = ~ 2 $Å^{-1}$.

A1.2.4 Prediction of theoretical anomalous differences

ASAXS experiments were simulated with CRYSOL v. 2.8.3 (158) from the atomic coordinates in PDB format for Tb^{3+} at 7517 eV, 7500 eV, and 7521 eV. Plots of solution scattering patterns covering a range from $0 \leq q \leq 2.0$ $Å^{-1}$ calculated to ensure observation of any weak larger q range scattering intensity differences. To account for the difference in scattering intensities for Tb^{3+} at the different x-ray energies, elements corresponding to the appropriate number of electrons scattering were substituted for Tb in the 2OJR coordinate set (table A1.1). Theoretical models of dLBT-OmpA, LBT-KRas, and dLBT-KRas were generated from splicing the LBT from 2OJR into loops in the OmpA (1BXW) and KRas-nanodisc (2MSC) coordinate sets.

A1.3 Results and Discussion

A1.3.1 Preliminary ASAXS measurements for Tb^{3+} loaded dLBT-ubiquitin

Towards the goal of ASAXS measurements from proteins embedded in nanodiscs, preliminary experiments were initially performed with a structurally well-characterized construct of ubiquitin with two engineered LBTs (dLBT-Ubq) (156). The expected SAXS and WAXS patterns for dLBT-Ubq were calculated with with CRYSOL (158) from the atomic coordinates for dLBT-Ubq (2OJR) with Tb^{3+} at energies remote and near to the Tb L_{III} x-ray abosorption edges (7517 eV, 7500 eV,

and 7521 eV). To account for the difference in scattering intensities for Tb^{3+} at the different x-ray energies, elements corresponding to the appropriate number of electrons scattered were substituted for Tb^{3+} in the 2OJR coordinate set (table A1.1).

Table A1.1. Scattering factors *f″* and *f′* for each x-ray energies remote and near to the Tb L_{III} absorption edge. Elements corresponding to the adjustment required to make number of anomalously scattering electrons correct for each energy tested for CRYSOL simulations.

Energy (eV)	λ (Å)	*f″*	*f′*	Equivalent # of electrons	Equivalent atom (atomic #)
7521	1.6485	11.65	-23.77	41.23	Nb (41)
7517	1.6494	6.02	-18.62	46.38	Pd (46)
7500	1.6431	4.16	-12.49	52.51	Te (52)

In solution, anomalous scattering is observable as differences between the scattering intensities observed in data collected at different energies. Simulated ASAXS experiments for dLBT-ubiquitin reveals differences in the scattering curves observed for intensities in the low- to mid-q region (~ 0.25 – 1.25 $Å^{-1}$) for the expected 7517 eV calculated data (figure A1.1A). The differences between the calculated intensities at remote, edge, and native intensity is at greatest 3.3% in this region of the scattering curve and less than 1% at $q < 0.2$ $Å^{-1}$ and $q > 1.2$ $Å^{-1}$ (figure A1.1B). Anomalous differences on this order of magnitude have been demonstrated to be experimentally measurable via ASAXS using iron anomalous centers in myoglobin with intensity differences between energies of

approximately 1% (155). Thus, theoretically, anomalous signal should be measurable for dLBT-ubiquitin.

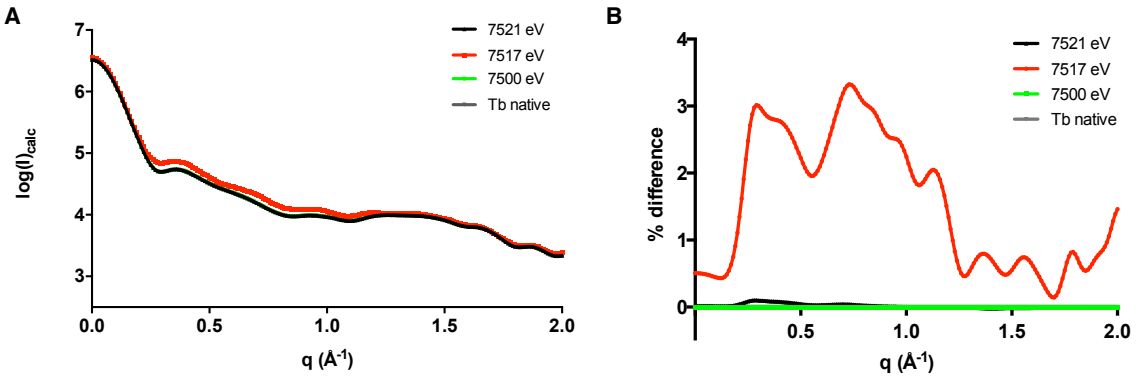

Figure A1.1. Expected results from dLBT-Ubiquitin ASAXS experiments. A, Scattering patterns calculated at the different energies used experimentally. Curves corresponding to 7521 eV, 7500 eV, and Tb native energies overlay. **B,** percent difference between the scattering curves at the different energies as a function of scattering angle.

Preliminary experimental ASAXS measurements of Tb^{3+}-loaded dLBT-Ubq were collected during commissioning at NSLS-II beamline 16-ID (LiX) at energies near (7521 eV and 7517 eV), and remote (7500 eV) to the absorption edge (L_{III} = 7514 eV) of the scattering atom (Tb^{3+}). Additionally, as a control, ASAXS data from the same dLBT-Ubq sample loaded with Ca^{2+} in place of Tb^{3+} was collected at the same energies (7521 eV and 7517 eV). Ca^{2+}, which binds tightly to the LBT (151), does not display significant scattering at these energies with f' = *0.336* e^- at 7521 eV and f' = 0.334 e^- at 7517 eV (159). Upon normalization of the data with the

scattering intensity peak from water, no observable differences were measured for Tb^{3+} anomalous scattering at the different energies tested (figure A1.2).

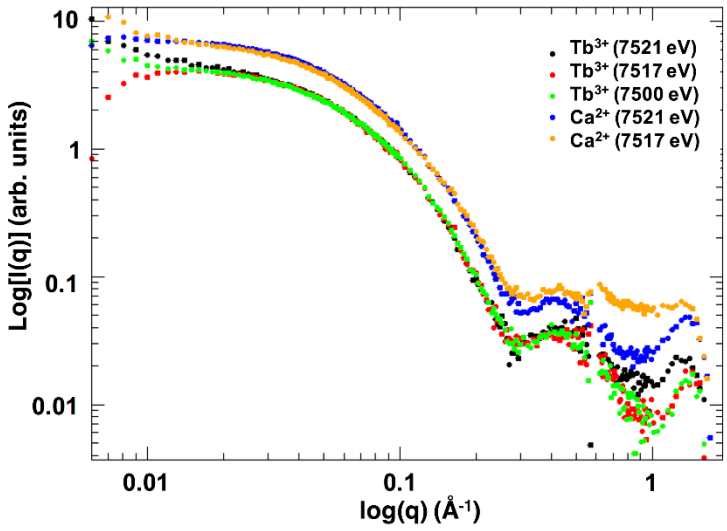

Figure A1.2. Double logarithmic plot of scattering intensity profiles for Tb^{3+}loaded dLBT-Ubq. Data collected at 7521 eV (black), 7517 eV (red), and 7500 eV (green) and Ca2+ loaded dLBT-Ubq collected at 7521 eV (blue), 7517 eV (orange).

The lack of anomalous difference measurability in the preliminary experiments could could be due to a number of different factors, including variation between the theoretical absorption edge and the experimentally determined absorption edge. Ideally, x-ray absorption spectra (XAS) of the element of interest would be acquired prior to ASAXS experiments to allow for appropriate beamline energy choice. However, during commissioning experiments at NSLS-II this capability was not available until after the start of

experiments. Thus, the theoretical absorption edge for Tb^{3+} was used as an approximation of the peak anomalous absorption (figure A1.3). It is evident from the XAS determined at the beamline (blue lines) that data collection for the above experiments was not performed at the true anomalous edge for Tb at the LiX beamline. Comparison of the scattering factor *f'* minimum for all lanthanides revealed there are only slight differences in the the number of anomalously scattering electrons across the lanthanides ($f' \cong 20\ e^-$) with the exception of La and Ce, which scatter to a less ($f' \cong 14\ e^-$). Thus, changing the lanthanide used in ASAXS experiments is not likely to have a significant effect on measurability of the signal. The greatest improvement in measurability is likely to come from optimized selection of experimental energies after performing an x-ray absorption scan at the beamline. Additionally, further optimization of sample quality, data collection parameters, and experimental geometry could lead to success in measurement of ASAXS intensity differences.

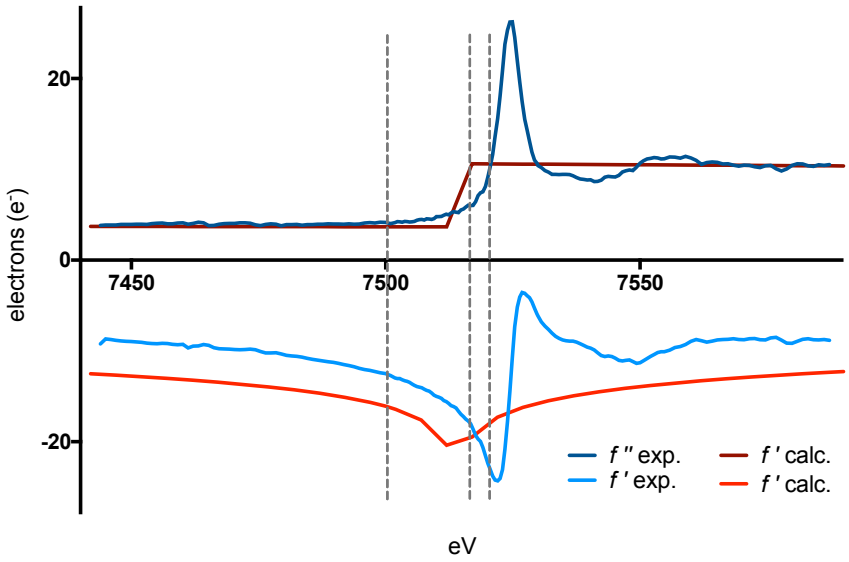

Figure A1.3. Theoretical (red) and experimentally-determined (blue) x-ray absorption spectra (XAS) plots in the region of the Tb L_{III} edge. Vertical dashed lines represent the energies at which ASAXS data was collected (7500 eV, 7517 eV, and 7521 eV). Experimental XAS plot created with CHOOCH(157). Theoretical *f′* and *f″* values based on the approximation by Cromer and Liberman (159) (accessed via http://skuld.bmsc.washington.edu/scatter/AS_periodic.html.)

A1.3.2 Effect of protein size on ASAXS difference intensity theoretical measurement with LBTs

Both proteins which have shown measured (155) or theoretically measurable ASAXS intensity differences, myoglobin and dLBT-ubiquitin, are relatively small proteins with molecular weights of 16.7 and 12.4 kDa, respectively. Tb^{3+}-loaded dLBT-Ubiquitin (111 aa) yields a 3% anomalous intensity difference compared to native with approximately 1 Tb^{3+} for every 50 amino acids contributing to the anomalous signal. Similarly, 1% difference ASAXS signal from

the intrinsic iron co-factor has been measured for myoglobin (154 aa) (155). Comparatively, the anomalous scattering from Tb stems from ~ 20 electrons versus ~ 7 electrons from iron; thus, the ratios of aa to anomalously scattering electrons to are 2o.9 e^- for myoglobin and 2.76 e^- for dLBT-ubiquitin. Therefore, it is conceivable that the increase anomalous signal from that arises from the additional anomalously scattering electrons in lanthanides (e.g. Tb) bound to LBTs engineered into proteins could provide sufficiently measurable signal from larger proteins and potentially more complex systems.

Simulation of ASAXS data via CRYSOL (158) allowed investigation of the effect of relative protein size on the measurability of ASAXS difference signal for two proteins larger than dLBT-ubiquitin and myogolbin: LBT-IL1β and dLBT-OmpA. Data simulated from the atomic coordinates for IL1-β with a single engineered LBT (PDB ID: 3LTQ) with Tb^{3+}-bound at 7517 eV, 7500 eV, and 7521 eV resulted in scattering curves which showed increased scattering intensity for data simulated at the Tb absorption edge (7517 eV) between q ≅ 0.25 $Å^{-1}$ and q= 0.6 $Å^{-1}$ (figure A1.4A). Dividing the scattering intensity profiles at remote and edge energies by the native data calculated the observable anomalous difference to be at most 1.4% of native in this region of the scattering curve and largely negligible (≤ 0.5% difference) for all other scattering angles (figure A1.4B). Thus, it appears

that with an approximately 3-fold increase in aa-to-scattering electron ratio observed for LBT-IL1β (8.37 e^-) in comparison to that of dLBT-ubiquitin, anomalous differences remain theoretically measurable.

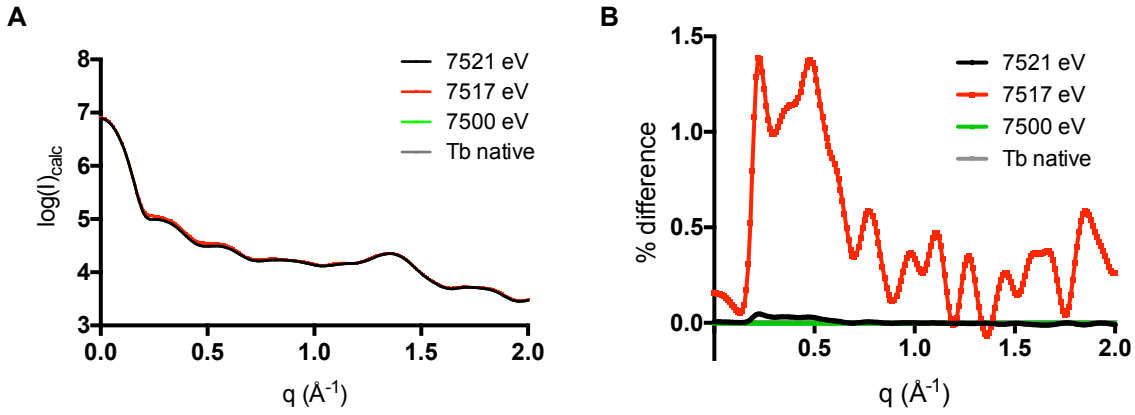

Figure A1.4. Predicted expected anomalous differences from Tb^{3+}-loaded LBT-IL1β. A, ASAXS intensity profiles calculated near and remote to the Tb absorption edge. Curves corresponding to 7521 eV (black), 7500 eV(green), and native (gray) energies overlaid. **B,** percent difference between the scattering curves at the different energies as a function of scattering angle. Curves corresponding to 7500 eV and native (gray) energies overlaid.

Furthermore, ASAXS data was simulated for a theoretical model of dLBT-OmpA (198 aa). The model of was created through insertion of single LBT tags at the N-terminus (periplasmic) and loop EL3 (extracellular) into a structure of OmpA from the PDB (PDB ID 1BWX). Single LBT tag structures were spliced from the experimentally-determined crystal structure of LBT-IL1β (PDB ID 3LTQ). Simulated ASAXS intensity profiles from q = 0 – 2 $Å^{-1}$ from theoretical atomic coordinates for dLBT-OmpA with Tb^{3+}-bound at 7521 eV, 7517 eV, and 7521 eV

(figure A1.5A). Minimal differences were observed between the scattering curves with the exception of slight increase in intensity corresponding to the data collected at 7517 eV b for the range ~0.25 < q < -0.5 $Å^{-1}$. Calculated anomalous differences between between the intensities at remote, edge, and native intensity is no greater than 0.5% of the native scattering intensity at any angle (figure A1.5B). Although the anomalous difference is minimal, the profile of the difference plot with a clear peak at ~0.25 < q < -0.5 followed by an oscillation about zero suggests the anomalous differences would be theoretically measurable within the at least within that q-range at the Tb Edge. However, a 0.5% difference may be exceedingly challenging to measure experimentally, and may be below the sensitivity of the practical ASAXS experiment on a whole. Additionally, it is worth noting that the aa-to-anomalous scattering electrons ratio is lower for dLBT-OmpA (4.86 e^-) than for LBT-IL1B (8.37 e^-); however, the apparent inverse-relationship between the ratio aa-to-anomalous scattering electrons and percent anomalous difference over native scattering does not hold in this case.

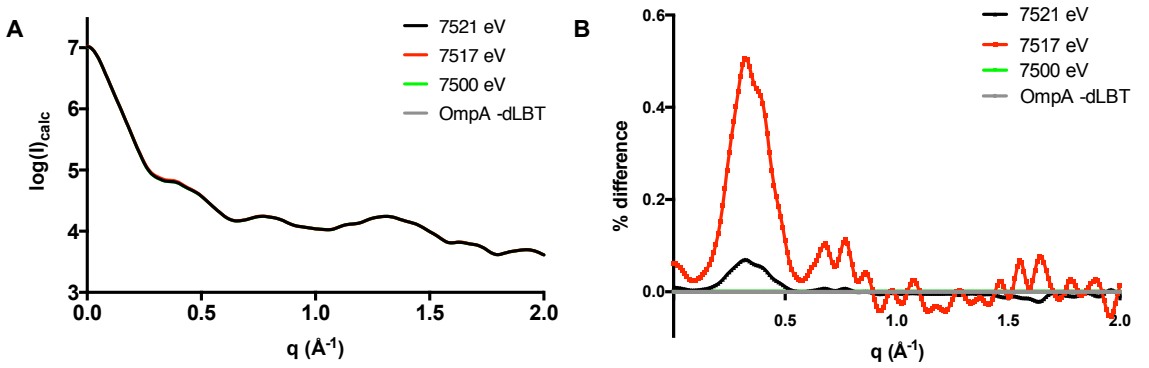

Figure A1.5. Predicted expected anomalous differences from Tb^{3+}-loaded dLBT-OmpA. A, ASAXS intensity profiles calculated near and remote to the Tb absorption edge. Curves corresponding to 7521 eV (black), 7500 eV (green), and native (gray) energies overlaid. **B,** percent difference between the scattering curves at the different energies as a function of scattering angle. Curves corresponding to 7500 eV and native (gray) energies overlaid.

A1.3.3 Predicted ASAXS signal from a protein tethered to a nanodisc

From the relatively simple test simulations of ASAXS data for dLBT-ubiquitin, LBT-IL1β, and dLBT-OmpA, Tb^{3+}- bound LBTs inserted into protein termini or loops enable sufficiently measurable signal, theoretically. Thus, to understand the feasibility of measurement of anomalous differences for significantly more complex systems, data from LBT and dLBT models of a protein resident on a MSP1D1 DOPC : DOPS nanodisc (ND) were simulated. The LBT- and dLBT-KRas · ND constructs were created by splicing the LBT (from 3LTQ) or dLBT (from 2OJR) into a loop on the KRas molecule distal from the ND in the KRas nanodisc model (PDB ID 2MSC) (figure A1.6A, D). Unsurprisingly, the simulated

ASAXS intensity profiles do not show any apparent significant difference between data simulated at different energies near or remote from the Tb absorption edge (figure A1.6B, C). The Tb^{3+} anomalous signal within the model is unable to be resolved from the total scattering of the resident KRas protein and MSP belts of the ND. Together the protein components of the ND contribute 601 aa; thus, the ratio aa-to-scattering electrons for this test system is 29.41 e^-. Adding an additional Tb^{3+} atom via dLBT-KRas does provide a significant improvement in the ratio of aa to scattering electrons for the ND constructs with 29.4 e^- for the single LBT and 14.9 e^- for the dLBT. However, the improvement is not reflected in percent difference for the anomalous scattering compared to native scattering, where the anomalous difference is only at most 0.3% of the native scattering at any given scattering angle (figures A1.6E and A1.6F). Ultimately, the anomalous difference signals form the anomalously scattering atoms in highly complex scattering systems, such as a resident-protein loaded into a ND, are likely too low for experimental measurement to be possible.

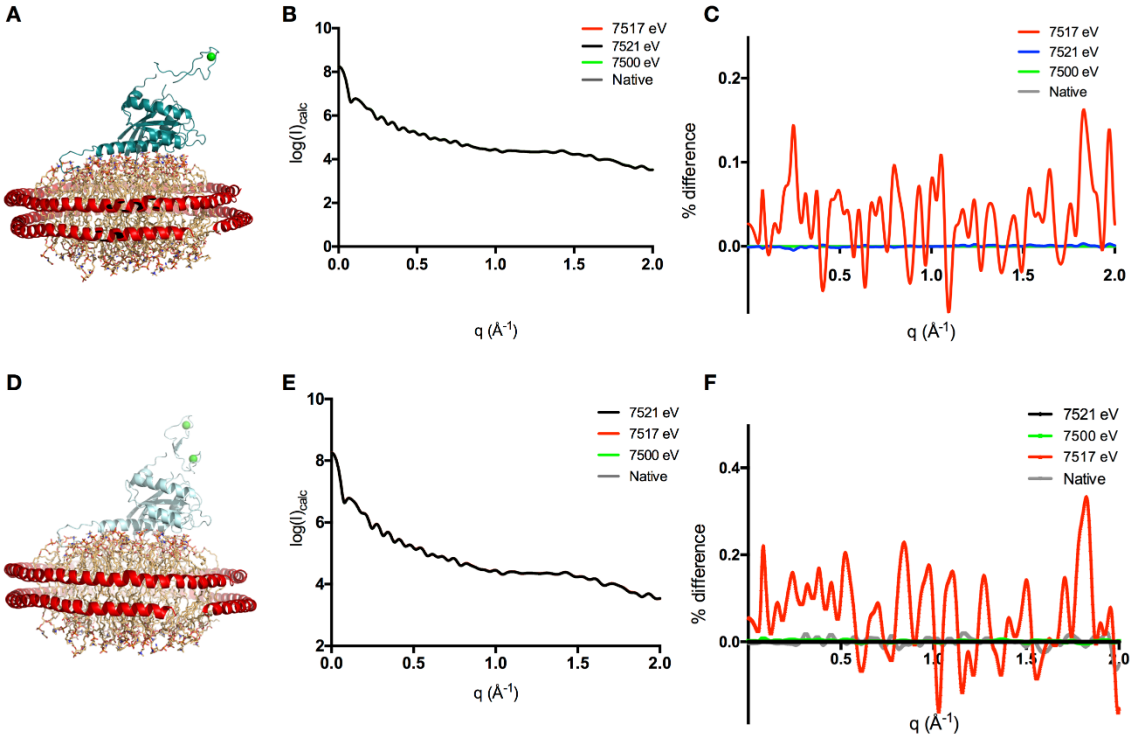

Figure A1.6. Simulated ASAXS data from Tb^{3+} -loaded LBTs on a protein tethered to a nanodiscs. A, Atomistic model of KRas with theoretical Tb^{3+} -loaded LBT spliced into a loop region tethered to MSP1D1 (80 : 20 DOPC:DOPS) nanodiscs, based on 2MSC. Tb^{3+} shown as green sphere. **B,** ASAXS intensity profiles calculated near and remote to the Tb absorption edge. Curves corresponding to 7521 eV (black), 7521 eV (red), 7500 eV(green), and native (gray) energies overlaid. **C,** percent difference between the scattering curves at the different energies as a function of scattering angle. Curves corresponding to 7521 eV (black), 7500 eV (green), and native (gray) energies overlaid. **D,** Atomistic model of KRas with theoretical Tb^{3+} -loaded dLBT spliced into a loop region tethered to MSP1D1 (80 : 20 DOPC:DOPS) nanodiscs. Tb^{3+} shown as green spheres. **E,** ASAXS intensity profiles calculated near and remote to the Tb absorption edge. Curves corresponding to 7521 eV (black), 7521 eV (red), 7500 eV(green), and native (gray) energies overlaid. **F,** percent difference between the scattering curves at the different energies as a function of scattering angle. Curves corresponding to 7521 eV (black), 7500 eV (green), and native (gray) energies overlaid.

A1.4 Conclusion and Future Directions

Overall, the work presented in this appendix found that measuring anomalous scattering intensity differences from preliminary ASAXS experiments with a structurally well-characterized construct of ubiquitin with two genetically-encoded LBTs will require further optimization of experimental design, geometry, and sample-preparation in order to measure the ~3% anomalous difference predicted for simulated ASAXS experiments by CRYSOL. Computational simulations of LBT-enabled ASAXS experiments to understand the feasibility of measurement of anomalous scattering intensity differences for both simple proteins and more complex systems revealed that the theoretically achievable ASAXS intensity differences proceeded were maximized in small globular proteins with a low aa-to-scattering electron ratio (table A1.2). However, the measurement of the theoretical differences (0.5%) predicted for a slightly larger protein system, OmpA, are likely outside of the current sensitivity of the practical experiment. Similarly, simulated data for a complex system of a single or double LBT-tag inserted into a protein tethered to a nanodiscs revealed negligible anomalous signal over all scattering angles.

Table A1.2. Theoretical ASAXS anomalous difference possible.

Sample	# aa	Scattering Atom (#)	aa-to- scattering electrons ratio	Theoretical Max % Diff.
Myoglobin	154	Fe (1)	20.9	1*
dLBT-Ubiquitin	111	Tb (2)	2.72	3.3
LBT-IL1β	171	Tb (1)	8.37	1.5
dLBT-OmpA	198	Tb (2)	4.86	0.5
LBT-KRas · ND	601	Tb (1)	29.41	0.15
dLBT-Kras · ND	612	Tb (2)	14.9	0.3

* Measured experimentally (155). Note: f'_{Fe}= -7.36, f'_{Tb}= -20.43.

Thus, as scientific questions are often asking increasingly complex questions of increasingly complex systems, it is unlikely that ASAXS will be able to provide sufficient anomalous to answer those questions. Ultimately, ASAXS, though still incredibly challenging, is better suited to relatively small proteins like myoglobin and ubiquitin. However, it would be interesting, in the future, to try to understand the relationship between the ratio of the number of amino acids in a protein to the number of scattering electrons and the theoretically achievable anomalous difference signal. From the very small test set presented in this appendix, a predictable inverse-relationship emerged. However, the model of dLBT-ubiquitin appeared to be an outlier where greater signal would be intuitively expected than was simulated from the aa-to- scattering electron ratio.

APPENDIX A2

Towards an understanding of the structural determinants of lipin1 cellular localization and function

A2.1. Introduction

Disturbances in processes contributing to lipid homeostasis have long been linked to the development of pathologic metabolic dysregulations, such as type-2 diabetes, metabolic syndrome, and obesity related dyslipidemia. Not only are these health concerns widely prevalent in today's population, their incidence and severity is continuing to increase at an alarming rate. Notably, loss-of-function mutations in the protein lipin-1 have been found to promote abnormal adipocyte differentiation, peripheral neuropathy, impaired glucose tolerance, circulating hyperlipidemia, and neonatal liver steatosis (160), all of which are hallmarks of dysregulated lipid homeostasis. Additionally, lipin1 has been shown to be directly phosphorylated in response to insulin signaling in humans, further corroborating the ties between type-2 diabetes and dysfunction of lipin-1 (161, 162). Expression of the lipin1 gene is highly regulated via hormonal signaling with lipin1 transcription being induced by fasting, high-fat diets, diabetes, chronic alcohol consumption, statins, and glucocorticoids (163).

Lipin-1, a member of the lipin family of proteins, is the predominant lipin isoform present in organs and tissues with high rates of lipid flux, including the liver, which plays a critical role in lipid metabolism and homeostasis in humans (164). Recently, lipin-1 has already been identified as a therapeutic target for disorders involving dysregulation of lipid homeostasis, dyslipidemia, and nonalcoholic fatty acid liver diseases and alcoholic liver disease (165). In addition to lipin-1 expression in the most impactful tissues for lipid homeostasis, it is poised as a unique target to attempt to quell dysregulation of lipid homeostasis as it plays a role in two functionally different pathways of lipid metabolism and homeostasis. Unlike most proteins, which play only a single role within the cell, lipin-1 displays a unique bi-functionality as a phosphatase and a transcriptional co-activator bridging glycerolipid biosynthesis and lipoprotein metabolism (figure A2.1).

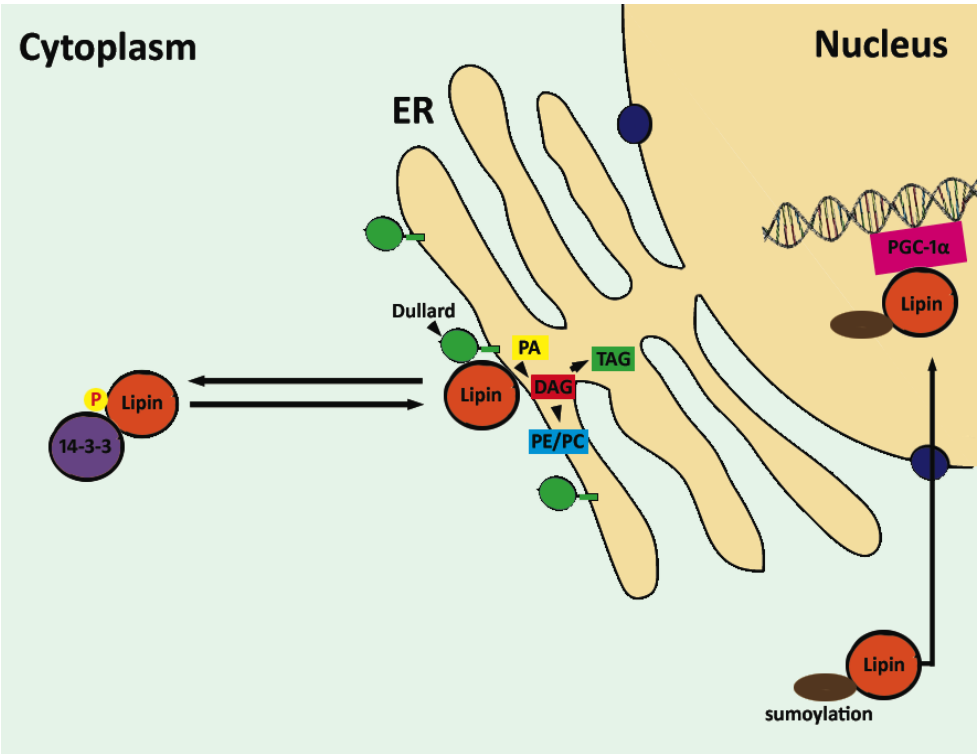

Figure A2.1. Schematic of the dual roles of lipin1 in the different cellular compartments cytoplasm (green), membrane, and nuclear (yellow). Figure courtesy of Dr. Jeremiah Farelli.

Functionally, lipin-1 plays a central role in triacylglycerol biosynthesis via its phosphatidic acid (PA) phosphatase activity through its catalysis of the conversion of membrane PA to diacylglycerol (DAG) (166). Additionally, lipin-1 acts as a transcriptional co-activator to increase the transcription of lipid oxidative metabolism and membrane biogenesis (165). Not only are the dual functions of lipin-1 mechanistically independent, they have been shown to be highly correlated to the localization of the protein within the cell. When localized to the plasma

membrane, lipin-1 carries out its PA phosphatase activity, and when localized to the nucleus lipin1 acts as a transcription co-activator via interactions with the transcription factors PPARα and PGC-1 (165). This localization of lipin1 is highly regulated by many different processes including covalent post-translational modifications, as well as non-covalent modulations by interactions with its binding partners.

Structurally, lipin-1 is a large complex molecule (890 aa) made up of multiple domains, known recognition motifs, and post-translational modification sites. The primary structure of lipin is compartmentalized by its different domains with a short amphipathic helix (aa 1-20) and a highly conserved NLIP (N-terminal domain of lipin) domain at the N-terminus (aa 21-106) of the protein and a haloalkanoic acid dehalogenase (HAD) phosphatase domain (aa 654-890) at the C-terminus (Figure A2.2). The domains at the termini are well conserved evolutionarily and have been predicted to assume well-characterized folds. The NLIP domain has been predicted via homology modeling to take on an immunoglobulin-like beta-sandwich fold, while the HAD domain takes on the characteristic HAD-domain rossmanoid core fold (167). Structural similarity is high, even when sequence similarity is low, among all HADSF family members, especially in the core domain, making homology modeling highly reliable. While

the domains at the termini are well conserved and therefore predicted to be well ordered, the sequence bridging the gap between the two termini is a 540 aa stretch of largely non-phylogenetically conserved, likely disordered sequence referred to as the intervening sequence. While the intervening sequence is thought to be largely disordered, it does contain important sequence elements including many of the sites for post-translational modification (phosphorylation and sumoylation), as well as recognition motifs for 14-3-3 proteins. In order to define the mechanism of lipin-1 function in the cell, information is needed about the basic structure-function relationship of lipin1 domains and sequence motifs (168, 169).

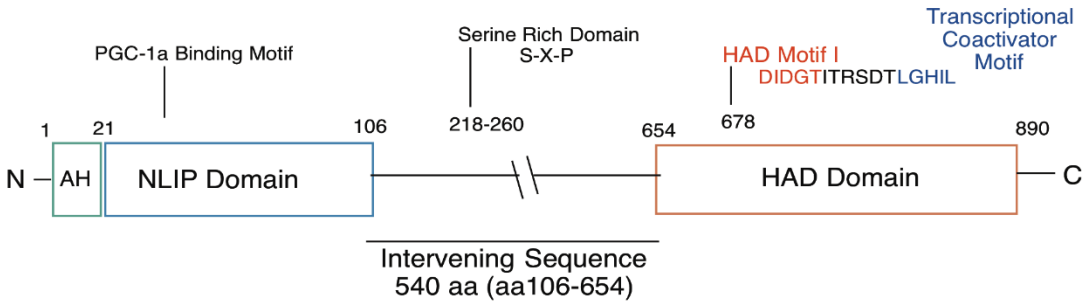

Figure A2.2. Domain structure of human lipin-1α. Boxed domains indicated predicted well-ordered domains.

While there are predictions for the structure of the well-ordered domains, there have been no experimental structure determinations for lipin-1. Any structural and biophysical information gained about lipin-1 will lead to a

fundamental understanding of the structural basis of the specific molecular interactions between lipin1 and its substrate, the membrane, and its binding partners, 14-3-3 and transcription co-activators, PPARα and PGC-1α. Ultimately this will help to define the poorly understood regulation of a key protein central to two pathways maintaining lipid homeostasis. Because of difficulties encountered previously in the Allen Lab with expression and purification of lipin-1 constructs, focus has shifted to expressing and purifying a homolog of lipin1 from *E. coli*, Nir2, and ortholog of lipin-1 from *S. cerviseae*, Smp2.

Nir2 shares 14% sequence identity (21% highly similar) with lipin-1. The construct of Nir2 used in the studies presented in this chapter is 345 amino acids comprising amino acids 911-1244 of the native Nir2 sequence. Functionally, Nir2 is a phosphatidylinositol-transfer and phosphatidic-acid binding protein (170). The construct employed in this chapter represents the two lipin-like domains and excluded the phosphatidylinositol transferase domain (figure A2.3A and B). Notably, Nir2 shares the DXDGT/V motif and Rossmanoid-core fold known in HAD superfamily members to be capable of phosphoryl-transfer (95). However, the catalytic Asp residue in Nir2 is mutated to serine giving rise to phosphatidic acid binding function over phosphatidic-acid phosphatase function as seen in lipin-1 (170). Similarly, to lipin-1 Nir2 has been demonstrated to translocate to the

plasma membrane in response to external signaling. Similarly, Smp2 shares high sequence similarity and domain structure to human lipin-1 (171, 172). Specifically, the construct employed in the studies presented in this appendix was a truncated construct which fused the NLIP and HAD-domains together (figure A2.3 C and B). Ultimately, by fusing the two well-ordered domains of Smp2 together, while omitting the intrinsically disordered intervening domain should allow for increased success in protein expression and purification over human lipin-1.

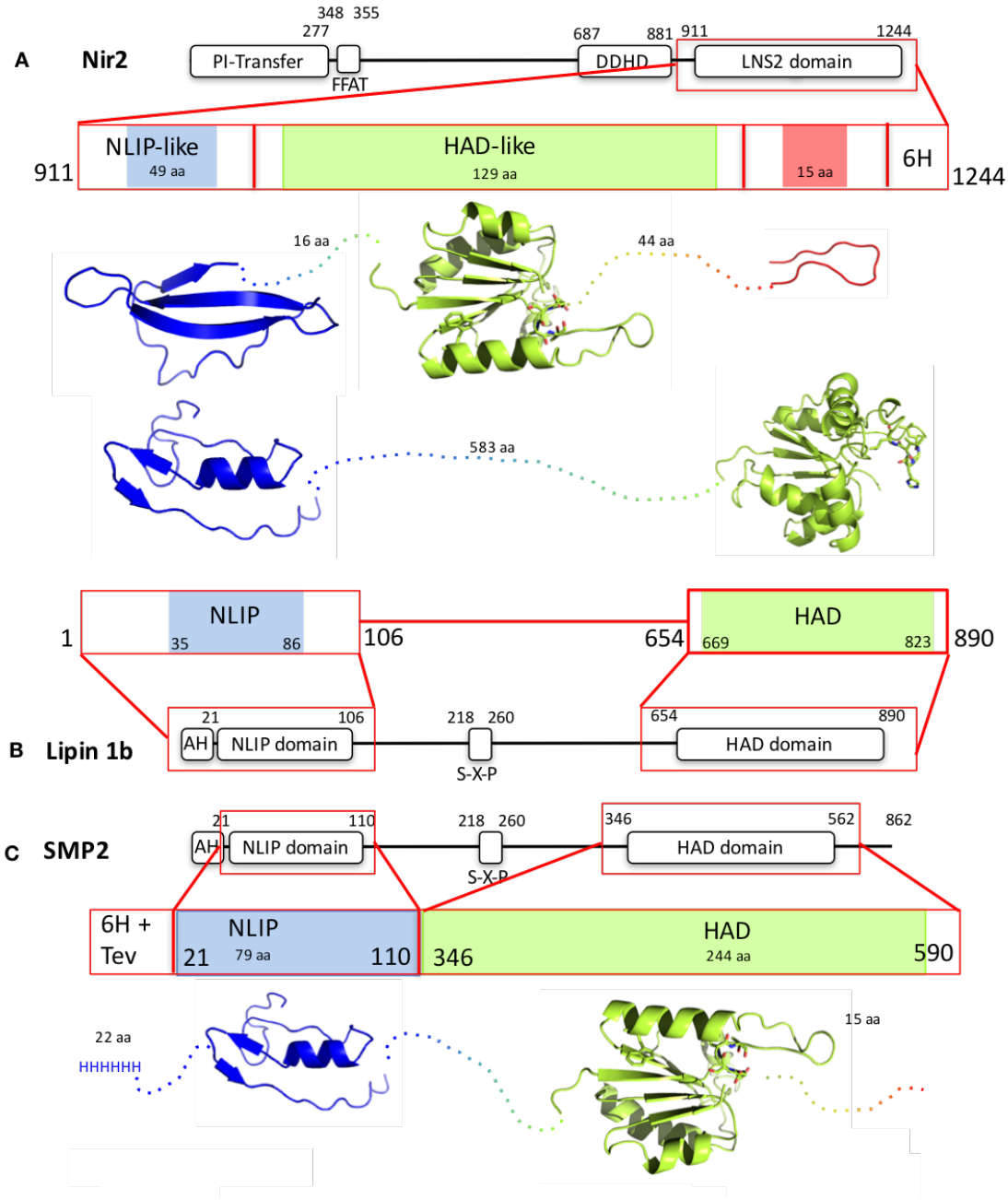

Figure A2.3. Comparison of Nir2, Smp2, and Lipin-1 domain structure. NLIP domains represented in blue, HAD domains represented in green, and highly-conserved Nir C-terminal domain represented in red. **A,** Domain structure of lipin-1 human homolog Nir2. **B,** Domain structure of Lipin-1β. **C,** Domain structure of lipin-1 yeast ortholog Smp2. Cartoon representations of the molecular structures of the individual domains created by Phyre2homology modelling (167).

Leveraging biophysical and structural biology tools, a quantitative and molecular view of the key mechanisms for lipin1 functional regulation can be obtained to corroborate insights gained from previous macro-level studies of the protein. Using a multi-faceted approach that combines genetic engineering of lipin1 constructs, *in vitro* thermodynamic and structural analyses, and biophysical techniques, accurate lipin1 models can be generated, which define the structure-function relationship for lipin1 domains and its molecular interactions with its substrate, the membrane, and its protein partners.

A2.2. Materials and Methods

Work on the expression, purification, and crystallization of the truncated Smp2 construct was performed in collaboration with undergraduate research assistant Merissa Brousseau.

A2.2.1. Constructs of Nir2, lipin, and Smp2

Table A2.1. Constructs of Nir2, Smp2, and lipin-1β used in this appendix.

Construct	Native Protein AA	Vector	Tag?	Antibiotic	MW (Da)	Experiment
Nir2 (911-1244)	911-1244	pET-24a	6His (non-cleav.)	Carb./ Cam.	37376.5	Crystallography SAXS
Truncated Smp2	21-100, 346-590	p15 (TEV)	6His	Carb./ Cam.	41961.5	Crystallography
Lipin-1*β* (human)	1-890	**	FLAG	**	98701.2	SAXS AUC MST

** human lipin-1β was purified from HeLa cells by James Eaton (Harris Lab, University of Virginia).

A2.2.2. Purification of Nir2 (911-1244)

One mL of overnight culture was inoculated into 1 L of Luria Broth media treated with 1x carbenicillin and 1x chloramphenicol. For each prep of Nir2, 4-6 L of media were inoculated. Cells were allowed to grow at 37 °C with shaking until an OD600 = 0.6-0.8 was reached. Once the appropriate OD was reached, cells were allowed to temperature equilibrate to 16 ° C prior to induction of Nir2 protein expression for at least 1 hour. Following temperature equilibration 0.5 mM IPTG was used to induce protein expression and cells were allowed to continue to grow with shaking overnight. The resulting cell suspensions were pelleted by centrifugation at 6,000 rpm. Cell pellets resuspended and homogenized in lysis buffer (25 mm HEPES pH 7.5, 500 mM NaCl, 10 mM Imidazole, 5 mM MgCl2, 2%

glycerol) with the addition of protease inhibitors and DNAse. Following homogenization cells were lysed with a microfluidizer and the lysate was pelleted in the ultracentrifuge at 38,000 krpm for 30 minutes at 4 °C. Purification of protein was achieved through Ni2+-NTA affinity chromatography using a peristaltic pump and commercially procured pre-packed HisTrap-HP column (GE Healthcare). Elution of Nir2 from the affinity column was accomplished with a step-gradient of 10%, 20%, 25%, 35%, 50%, 75% and 100% elution buffer (25 mM HEPES pH 7.5, 150 mM NaCl, 500 mM imidazole, 5 mM MgCl2, 2% glycerol) over 20 column volumes. Nir2 Purification from 6 L of inoculated culture generally yields 30 mL of 0.5 mg/mL pure Nir2 protein. Pure protein was dialyzed overnight against 4 L of protein buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 5% glycerol) to remove excess imidazole. Following dialysis any remaining precipitation was removed by centrifugation at 38,000 rpm for 30 minutes at 4 °C. Homogeneity of the resulting pure protein was characterized by dynamic light scattering (DLS) at 4 °C.

A2.2.3. Differential scanning fluorimetry (Thermofluor)

To screen for optimized buffer conditions via the Thermofluor Solubility Screen (Hampton Research) , 1000 μL of 20 μM protein (Nir2 or Smp2) with 2 μL

of Sypro Orange Dye (500X, from Invitrogen) was prepared by dilution to the with water. To each well of a 96 well white-walled PCR tray (Eppendorf), 10 μL of protein solution plus 10 μL of well solution were added for a final reaction volume of 20 μL. Clear sealing fil was applied to the top of the plate to mitigate evaporation of sample during the the thermal melt assay. The fully prepared assay plate was placed onto the heat-block of a RealPlex ep Mastercycler thermocyler (Eppendorf) and the lid was closed. The protein samples were subjected to a termal melt program that starts at a constant 4 ° C step for 15. Following temperature equilibration at 4 °C, the temperature is increase at 1°C per minute until the temperature reaches 95°C. After the temperature ramp is completed, the temperature is again held constant at the high temperature (95 °C). During the assay, fluorescence at 520 nm is monitored.

A2.2.4. Purification of truncated yeast SMP2

Truncated yeast Smp2 was overexpressed in *E. coli* (DE3) with overnight cultures diluted 1:1000 in YT medium containing 100 μg/ml carbenicillin and 35 μg/ml chloramphenicol. Cultures were agitated continuously at 225 rpm at 37° C to an OD_{600} of 0.6 at which point expression of Smp2 was induced by addition of IPTG to a final concentration of 1 mM. Cultures then continued growing overnight

with agitation at 16° C. Cells were then pelleted at 6,000 RPM for 15 and frozen at -80°C. Cells were suspended in lysis buffer (50 mM BisTris Propane (pH 6.2), 500 mM NaCl, and 20 mM imidazole) and lysed by microfluidization at 18,000 psi. The lysate was centrifuged at 38,000 RPM for 30 min. Clarified lysate was loaded onto an ÄKTAxpress FPLC (GE Healthcare). Lysate was passed through a 5 mL HisTrap FF column (GE Healthcare), washed with 10 column volumes of lysis buffer, and eluted in buffer containing 50 mM BisTris Propane (pH 6.2), 100 mM NaCl, and 500 mM imidazole. Peak fractions were collected and protein was analyzed by SDS/PAGE, and run through an additional round of column chromatography.

A2.2.5. Expression testing of truncated Smp2

To test for optimal induction conditions of the truncated construct of Smp2, twelve colonies of BL21 cells were used to inoculate three groups of LB media, each of which was grown to an OD of 0.4 then two groups were induced by addition of IPTG to a final concentration of 1 mM. One group was grown at 37° C for 3 h post-induction, one group was grown at 16° C overnight, and one group was grown in minimal media at 37°C for 12 hours to test for auto-induction. The auto-induction cells failed to grow.

To test for optimized expression across differing cell-lines, a pUC vector containing the gene for Smp2 was used to transform BL21-CodonPlus (DE3) Competent Cells, Arctic Express (DE3) Competent Cells, BL21-Gold(DE3)pLysS Competent Cells, and SoluBL21 (DE3) Competent Cells from Agilent. All competent cells were thawed on ice for 10 minutes. XL10-Gold β-mercaptoethanol mix was diluted 1:10 in de-ionized water and 2 μL were added to the CodonPlus and Arctic Express cells during incubation on ice for 10 minutes with slight agitation every 2 minutes. Each cell line was transformed with 5 μL containing 1-50 ng of plasmid DNA and incubated on ice for 30 minutes. The cell lines were then heat shocked at 42° C for 20 seconds and incubated on ice for 2 minutes. Then 0.9 mL of preheated (42 °C) SOC medium was added to each transformation reaction and the reactions were incubated at 37 °C for 1 hour with shaking at 225–250 rpm. Finally, the reactions were plated on plates containing 1x chloramphenicol and grown overnight at 37°C. The next day, three colonies were selected from each cell line and used to prepare and overnight culture. The overnights were each grown at 37° C to an OD_{600} of 0.4 and IPTG was added to a final concentration of 1mM. The cells were allowed to continue growing at 37° C for 3 h post-induction. Pre- and Post-induction samples were collected and compared via gel electrophoresis.

To compare the effects of final concentration of IPTG on expression, three colonies of CodonPlus cells were used to inoculate overnight cultures of LB. The overnights were each grown at 37° C to an OD_{600} of 0.4 and IPTG was added to final concentrations of 0.05 mM, 0.2 mM, 0.5 mM and 1.5 mM IPTG. The cells were allowed to continue growing at 37° C for 3 h post-induction. Pre and Post-induction samples were collected and compared via gel electrophoresis. Finally, the addition of a Rosetta (pRare) plasmid was tested both through co-transformation with the Smp2 plasmid and with a 2 step transformation in which competency was reintroduced after transformation with the Rosetta plasmid. The transformation was performed as above with DNA containing the Rosetta plasmid, the following day, a single colony of transformed cells was used to inoculate 10 mL of LB containing 1x antibiotic and was grown overnight at 37° C shaking at 225–250 rpm. The next day, 2 mL of this culture was added to 100 mL fresh LB and grown to an OD_{600} of 0.6. The cells were then poured into 50 mL conical tubes and incubated on ice for 10 minutes. Next the cells were centrifuged at 6000 RPM for 15 min, the media was removed and each tube was re-suspended 10 mL 0.1 M CaCl2. The two tubes were combined then pelleted again and the $CaCl_2$ was poured off. The final cell pellet was re-suspended in 4 mL 0.1 M $CaCl_2$ and 15% glycerol and left on ice, overnight in a 4° cold room.

A sample of 50 μL of the resulting suspension was aliquoted into 1.5 mL centrifuge tubes were frozen in liquid nitrogen and stored at -80 °C. One aliquot, however, was used in a second transformed with Smp2 plasmid DNA as described above. A single colony of transformants were used to inoculate 10 mL of LB containing 100 μg/ml carbenicillin and 35 μg/ml chloramphenicol and was grown overnight at 37° C shaking at 225–250 rpm. This starter culture was used to inoculate fresh LB the following day which was grown to an OD_{600} of 0.4 and IPTG was added to final concentrations of 1 mM IPTG. The cells were allowed to continue growing at 37° C for 3 hours post-induction. Pre -and Post-induction samples were collected and compared via gel electrophoresis.

A2.2.6. Crystallization of Nir2 and truncated yeast Smp2

Crystallization condition screening of purified Nir2 protein was carried out via commercially-available high-throughput (HT) sparse-matrix screens (96 well) from Hampton Research and Microlytic. Concentration of protein (7 - 23 mg/mL), temperature (4 °C, 17 °C, 25 °C, and 37 °C), and crystallization experiment geometry (hanging-drop, sitting-drop, and under-oil) was varied for high-throughput screening. Conditions producing crystal hits, confirmed by IZIT dye (Hampton Research) (173) or diffraction screening, were further optimized

through grid screening focused on variation sin protein concentration, pH, precipitant concentration, protein: well solution ratio, or slowed equilibrium in a 24-well hanging-drop crystallization geometry.

HT sparse-matrix crystallization condition screens were prepared with purified Smp2 protein concentrated to 15 mg/ml in 96-well sitting-drop geometry. Individual crystallization drops were prepared with 1 μl of protein and 1 μl of screening condition. Following set up at room temperature (~25 ° C), crystallization plates were incubated at 17 °C for crystal growth. Crystal hits identified through HT screens were optimized via similar grid-screening optimization strategies employed for Nir2.

A2.2.7. SAXS data collection and data processing for Nir2 and human lipin-1β

Solution SAXS experiments of Lipin-1β and Nir2 were performed at X9 at Brookhaven National Lab (BNL) National Synchrotron Lightsource (NSLS-I) (Upton, NY). Nir2 and lipin-1β samples, prepared as described above, were transported to the beamline on wet ice. Prior to experiments, all samples were centrifuged for at least 30 minutes at 18,000 rpm (rotations per minute) on a table top microcentrifuge followed by filtration to remove any aggregates within the sample. The Nir2 samples were analyzed over a concentration range of 1.3, 1.9,

2.75, and 3.98 mg/mL and with reserved sample dialysis for each sample serving as the buffer blank for subtraction. Lipin-1β samples were run at two concentrations 1 mg/mL and 0.5 mg/mL. For each data set, 20 μL of sample was was drawn into the experimental capillary. Images were collected with a 30 second exposure at λ = 0.98 Å were collected on a MarcCCD detector positioned 3.4 m from the sample resulting in a resolution range of $0.0064 < q = 4\pi\sin\theta/\lambda > 2.0$ Å^{-1}, where 2θ is the scattering angle.

Data images were scaled and radially integrated via pyXS scripts implemented at the beamline. Scattering profiles computed for each image collected for the sample (n = 3) were subsequently averaged. The buffer contribution to the sample scattering was removed by subtraction of the scatting from the matched buffer sample form the sample to give the one-dimensional scattering intensity profile for the protein samples. Further processing and analysis of the SAXS data was carried out with the ATSAS suite of programs (132). Radius of gyration (R_g) for each sample was determined from transformation of the scattering intensity profiles in PRIMUS (133) via the Guinier approximation (126) within a fitting limit of $qR_g \leq 1.3$. The particle distiance distribution functions, P(r), for each sample were calculated from the scattering intensity curves in PRIMUS using the GNOM and autoGNOM tools to determine the longest inter-

particle vector, D_{max}. The molecular envelopes of lipin and Nir2 were generated by calculation of 4 *ab inito* shape reconstructions in DAMMIN that were then averaged by DAMAVER to select the most populated volume for all reconstructions.

A2.2.8. AUC data collection and data processing for human lipin-1β

Human Lipin-1β with a C-terminal FLAG-tag purified from HeLa cells was provided for sedimentation equilibrium (SE) and sedimentation velocity (SV) AUC experiments by James Eaton (Harris Lab, University of Virginia). Absorbance optics SV-and SE-AUC) experiments using a Beckman XL-A/I ultracentrifuge, equipped with a Ti50An rotor using a 12mm six channel centerpieces for SE and 12mm two channel centerpiece for SV with quartz windows and detection by UV at 280 nm. AUC experiments out at 4 °C in a buffer of 20mM HEPES pH 7.2, 150mM NaCl, and 10% glycerol. For SE, protein at concentrations of 0.7, 0.5, and 0.2mg/mL Samples were subjected to spins at 6400 rpm for 21 hours to achieve equilibrium before five replicates of two scans were recorded. Rotor speed was to 8000 rpm for 21 hours to achieve a new equilibrium before two scans were recorded. Rotor speed was increased to a final speed of 12000 rpm for an additional 21 hours to establish a new equilibrium before the

final 2 scans were recorded. For SV experiments, lipin-1β at a concentration of 0.6mg/mL was subjected to centrifugation at 42,000 rpm for 19 h allowing for collection of a total of 867 scans.

SV data was fit to a Lamm equation model of rapid-self-association equilibrium between monomer and dimer or monomer, dimer, and tetramer with the following parameters: $\bar{\nu}_{sample}$ = 0.729 cm^3 g^{-1}, ϱ_{buffer}= 1.035 g ml^{-1}, η_{buffer} = 0.015 P in SEDFIT (139, 174). Both models fit the data with rmsd = 0.008 – 0.004 between the model and experimental data. Notably, to achieve good fits to the data for both models, a term accounting for a small (s ≅ 0.1), rapidly sedimenting particle. SE data was initially processed in SEDFIT, and analyzed globally in SEDPHAT (139). Global reduced chi-squares (gcrs) indicating goodness of fit were between ~ 3-7 for the monomer-dimer-tetramer model for fitting.

A2.2.9. Microscale thermophoresis (MST)

Preliminary microscale thermophoresis experiments were performed using a Nanotemper Monotlith NT.115 insturment. Lipin-1β was fluorescently labeled on solvent exposed lysine residues with a 2x excess of fluorescent dye NT-495 dye (Nanotemper) resulting in labelling of an estimated 5 lysines/ lipin molecule. Labeled lipin (100 μM) held at a constant concentration of 1.8 μM was titrated

against increasing amounts of purified PPARα (100 μM, 26 μM) provided by Dr. Jeremiah Farelli (Allen Lab, BU) from 0.9 nM – 59 nM in 100 mM Tris buffer at pH 8.0. Thermophoretic movement of the lipin molecules was observed as a change in fluorescence signal. Fitting the fluorescence data to the Hill equation (175) allowed calculation of the equilibrium dissociation constant $K_d = 16 \pm 7$ nM.

Additionally, MST was used to probe for binding between full-length lipin-1β and phospholipid membrane bilayers (nanodiscs) *in vitro*. Fill length lipin-1β was labeled at lysine residues with a 5x excess of NT-495 dye. This fluorescent lipin-1β (100 uL, 1.8 μM) was then titrated with increasing amounts of membrane 8 nm nanodiscs (MSPΔH5 construct, DMPC/DMPG 3:1) with a concentration range of 0.08 nM- 11μM in 100 mM Tris buffer pH 8.0. Fluorescence data was fit to the Hill equation (175) to give an estimation of $K_d = 21$ nM.

A2.3. Results and Discussion

A2.3.1. Optimization of Nir2 and truncated Smp2 protein for structural studies

Previous work in the Allen lab towards an understanding of the unique structural underpinnings of cellular localization and function in lipin1 encountered difficulty with expression and purification of lipin-1 constructs primarily due to low expression in the *E. coli* expression system and high

aggregation propensities during purification. As a result, high-resolution structural studies by x-ray crystallography focused on a human homolog, Nir2, and yeast ortholog, Smp2, of lipin which allowed for more reliable expression and purification of the high-quality protein needed for crystallography.

Although it was possible to obtain pure protein protein of Nir2, the protein was largely unstable with high degrees of aggregation and precipitation seen in preparations of Nir2. As a result, low yields of protein of sufficient quality for structural studies was obtained from a 6 L prep. In attempt to mitigate Nir2 instability, differing purification and protein handling strategies were tested. Notably, adding a gel filtration step and low-pH elution strategies to the purification protocol had no effect in ameliorating aggregation during Nir2 preparation. To address the possibility that the aggregation propensity was coupled to temperature, dynamic light scattering (DLS) at different temperatures was performed. DLS experiments revealed Nir2 was extremely stable and monodisperse at 4°C and 17°C, but at 23°C, the entire sample displayed aggregation (figure A2.4A), thus, indicating that between 17°C and 23°C, the protein crosses a stability threshold and aggregates. Additionally, differential scanning fluorimetry (176) allowed for selection of optimal buffer conditions (50 mM HEPES pH 7.5, 100 mM NaCl, and 5% glycerol), which revealed a ~ 3 °C

increase in melting temperature ($T_{m,\ optimized}$ = 48.0 ° C) for for Nir2 in comparison to the original buffer (25 mm HEPES pH 7.5, 500 mM NaCl, 10 mM Imidazole, 5 mM $MgCl_2$, 2% glycerol, $T_{m,\ original}$ = 45.3 ° C) (figure A2.4B). Notably, the components of the original and optimal buffers for Nir2 are highly similar with the exception of the exclusion of imidazole. Unsurprisingly, imidazole has been shown in other his-tagged protein systems to lead to instability and insolubility (177). As a result of changing the buffer and maintaining sample temperature at 17 °C or below throughout protein purification, pure, stable Nir2 in sufficient quantities for high-throughput crystallization screening were obtained (figure A2.4C).

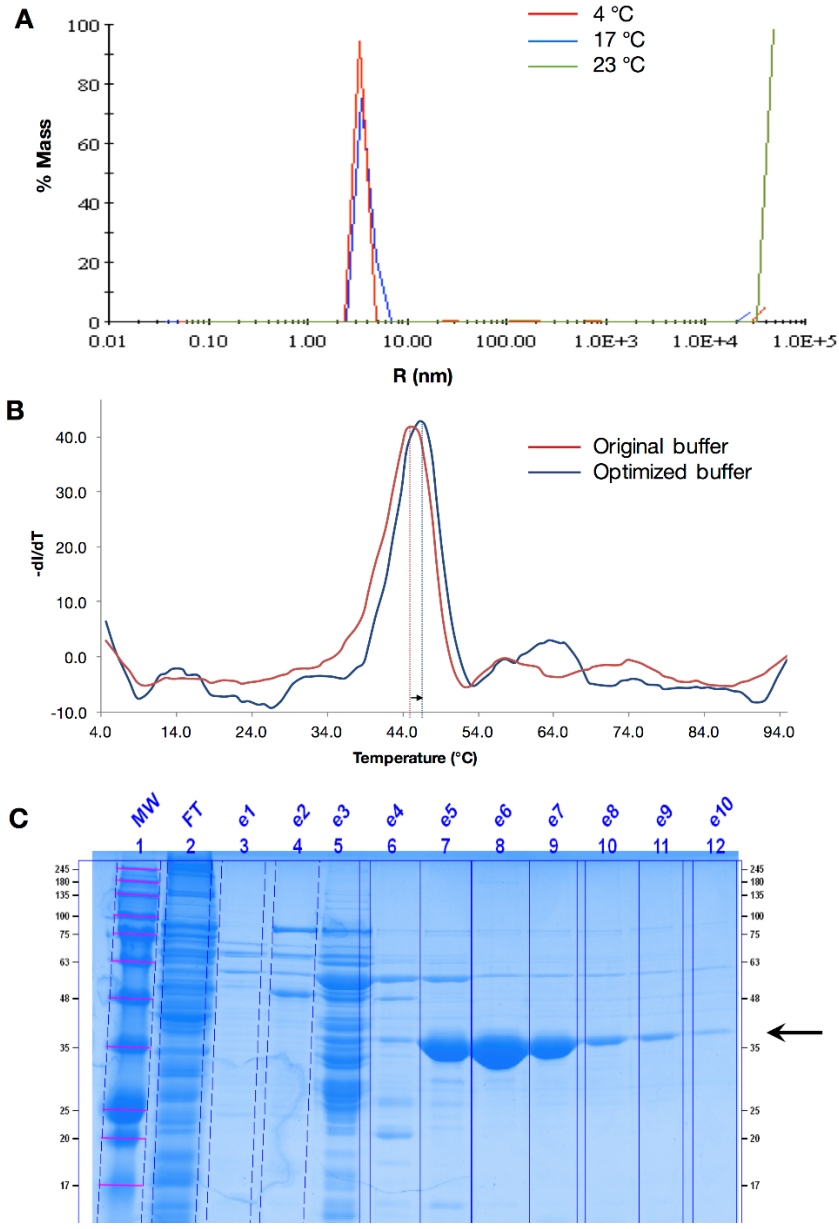

Figure A2.4. Optimization of Nir2 Purification. A, DLS traces of Nir2 at 4 °C (red trace), 17 °C (blue trace), and 23 °C (green trace). **B,** Thermal melting curves for Nir2 in the original purification elution buffer (25 mm HEPES pH 7.5, 500 mM NaCl, 10 mM Imidazole, 5 mM $MgCl_2$, 2% glycerol; $T_{m,\ original}$ = 45.3 ° C; red trace) and optimized buffer (50 mM HEPES pH 7.5, 100 mM NaCl, and 5% glycerol; $T_{m,\ optimized}$ = 48.0 ° C; blue trace). **C,** SDS-PAGE gel of purified Nir2 for x-ray crystallography studies. Bands corresponding to Nir2 indicated by the arrow.

In addition to purification of Nir2, purification and optimization of a truncated construct of Smp2, a homolog of lipin-1 from yeast was attempted for crystallographic studies in collaboration with undergraduate research assistant Merissa Brousseau. The truncated construct of Smp2 was reliably purified with high purity; however, low-levels of protein expression lead to overall yields of purified truncated Smp2 that were often too low to perform structural studies. As such, protein expression strategies, namely varying post-induction growth time and temperature, cell-line, and IPTG induction concentration, were assayed to determine an optimal expression system for truncated Smp2. Comparison of protein-expression levels following varied growth times and temperatures post-IPTG induction revealed no significant differences between cells allowed to grow at 37 °C for 3 h and cells allowed to grow for 16-18 h at 16 °C (Figure A2.5A). Transformation of truncated Smp2 into BL21-CodonPlus(DE3) competent cells, ArcticExpress(DE3) competent cells, BL21-Gold(DE3)-pLysS competent cells, and SoluBL21(DE3) competent cells from Agilent allowed for testing of the effect of cell-line on the protein expression levels of Smp2. BL21-CodonPlus(DE3) and ArticExpress(DE3) competent cells offered increased expression levels in comparison to the original soluBL21(DE3) cell line used (figure A2.5B). Both BL21-CodonPlus(DE3) and ArcticExpress(DE3) competent cells are specifically

engineered to aid in heterologous expression of non-bacterial proteins in *E. coli.* Specifically, BL21-CodonPlus(DE3) (Agilent) cells are specifically engineered with additional copies of tRNA genes to aid in heterologous expression in *E. coli* (178) and ArcticExpresss(DE3) (Agilent) cells aid in increasing protein folding through concurrent over-expression of specific chaperone proteins (179). As expression levels between the CodonPlus and ArcticExpress cell lines were equivalent, CodonPlus cells were selected for further expression of the truncated Smp2 due to the relative simplicity of growth protocol. Similarly, to the expression time and temperature, variation of IPTG had negligible effect on protein expression levels (figure A2.5D). Overall, increased levels of Smp2 protein expression were achieved through changing the cell line for heterologous expression of Smp2 in *E. coli.*

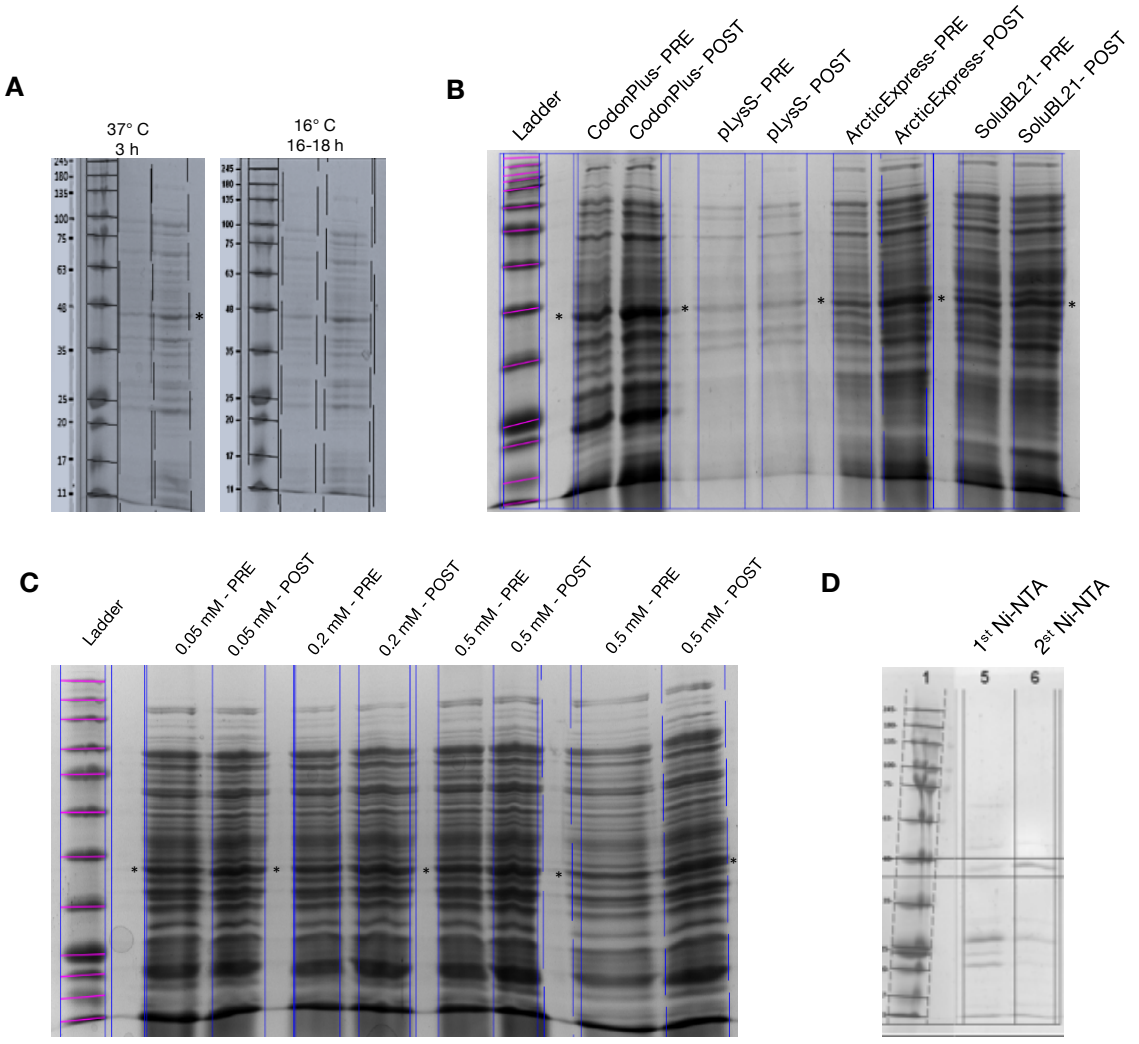

Figure A2.5. Results of expression testing for truncated Smp2 (indicated by *). **A,** Varied growth time and temperature. **B,** Varied cell lines. **C,** Varied induction concentration of IPTG. **D,** Final purification yield of truncated Smp2 after the first and second His-Trap purification steps. Gray horizontal bars identify the bands corresponding to Smp2. Experiments performed by undergraduate research assistant Merissa Brousseau.

A2.3.2. Crystallization of Nir2 and truncated Smp2 constructs

Optimization of purification and expression protocols for Nir2 and truncated Smp2 allowed for production of sufficient quantities of pure, monodisperse protein for structural studies by x-ray crystallography. High-throughput crystallization screening of Nir2 yielded twelve initial crystal hits (Table A2.2), which were then further screened for diffraction. Of these twelve initial hits, only two crystals did not diffract as salt; however, they did not diffract past 20 Å (figure A2.6). Optimization through grid screening around these condition was attempted, but in addition to no production of crystals of improved diffraction quality, the initial hit was not reproduced. Additionally, the level of precipitation observed in the high-throughput crystal screens (approximately 50%) was below ideal (80-90%), indicating the protein concentration used to set up crystals should be increased (173). Subsequently, increasing the protein concentration from 10 mg/mL, used in the initial screen, to 20 mg/mL for the second iteration of the screen, did not yield any positive crystal hits or reproduction of the initial hit.

Table A2.2. Initial Nir2 crystal hits from high-throughput crystallization screens.

Well	Condition	[Nir2] mg/mL	Drop ratio (prot.: well)	Izit Dye	Diffraction?
MCSG1 D11	0.2M Sodium Iodide 20% w/v PEG 3350	10.12	1:1	+	protein, >20 Å
MCSG2 F01	0.2M NaCl 0.1M Imidazole pH 8.0 0.1M Ammonium Phosphate Dibasic	10.12	1:1	-	SALT
MCSG2 C07	0.1M Tris:HCl pH 8.5 1.5M Ammonium Phosphate Dibasic	10.12	1:1	-	SALT
MCSG2 E06	0.1M Sodium Acetate pH 4.5 1M Ammonium Phosphate Dibasic	10.12	1:1	-	SALT
MCSG2 G06	0.2M NaCl 0.1M Sodium Citrate pH 5.5 1.0M Ammonium Phosphate Dibasic	10.12	1:1	-	SALT
MCSG3 A04	1M Imidazole pH 8.0 1M Ammonium Phosphate Dibasic	9.86	1:1	-	SALT
MCSG3 E01	0.2M NaCl 0.1 M NaH_2PO_4/K_2HPO_4pH 6.2 50% (v/v) PEG 200	9.86	1:1	-	SALT
MCSG4 C10	1 M NaH_2PO_4/K_2HPO_4 pH 8.2	9.86	1:1	-	SALT
MCSG2, A05	0.2M Lithium Sulfate 0.1 CAPS pH 10.5 1.2M NaH_2PO_4/K_2HPO_4	20.3			
MCSG4, A12	0.1M Bis-Tris Propane pH 7.0 4M Sodium Nitrate	20.3	1:1	-	SALT
PegRx HT	4 M Potassium Formate 0.1 M BisTris propane pH 9.0 2% w/v PEG monomethyl ether 2000	14.5	1:1	+	protein, >20 Å
MCSG3, A4	0.1M Imidazole, pH8.0 1M Ammonium Phosphate Dibasic	9.86	1:1	-	SALT

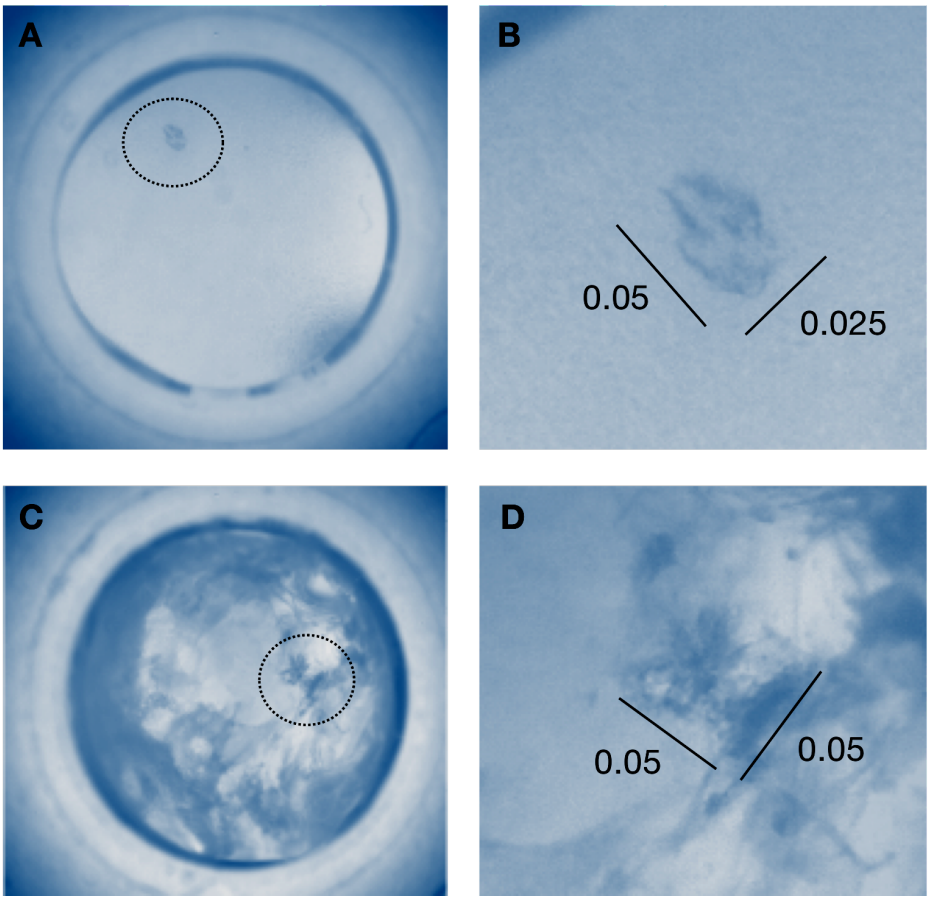

Figure A2.6. Initial crystal hits of Nir2 diffracting to > 20 Å. A, Crystal hit from 10.12 mg/mL Nir 2 in 0.2 M Sodium Iodide, 20% w/v PEG 3350. **B,** Zoom of circled region in A to show crystal detail. Dimensions given in mM. **C,** Crystal hit from 14.5 mg/mL Nir 2 in 4.0 M Potassium Formate, 0.1 M BisTris Propane pH 9.0, and 2% PEG monomethyl ether 2000. **D,** Zoom of circled region in C to show crystal detail. Dimensions given in mm.

Following the failure of continued grid-screening and HT-screening, *in situ* proteolysis via the Proti-ace screen (Hampton Research) was used to proteolytically cleave any flexible loops or domains that may be hindering crystallization (180). While overall, less heavy precipitate was observed in the

wells of the *in situ* proteolysis crystallization experiments, but no additional crystal hits or reproductions were observed. The use of additives can often help coax recalcitrant proteins into forming crystals (173). Using detergent (β-octyl gluocoside), the native substrate (PA) intrinsic co-factor (Mg^{2+}) of Nir2, and a crown ether observed to mediate protein crystal contacts (118) as additives to the Index HT screen (Hampton Research) did not produce any additional crystal hits. In continuing with crystallization screening for Nir2, it would be worthwhile to explore robotic based high-throughput crystallization at the Hauptman-Woodward Institute to allow screening of many more conditions with the limited quantities of crystallography-quality protein obtained (116). Additionally, using a crystal-former device to allow for the formation of concentration gradients and therefore increase the effective number of crystallization conditions screened (181).

High-throughput crystallization screening of Smp2 yielded four initial crystal hits (table A2.3). The crystals produced in the high-throughput 96-well trays were too small in size to be looped and screened for diffraction. However, dying with izit dye which penetrates into the solvent channels of protein crystals, but is excluded from inorganic salt crystals, revealed all hits with the exception of those from the $MgCl_s$ and PEG 3350 were likely protein crystals (figure A2.7). Attempts to reproduce and optimize these initial crystal hits to produce

diffraction-quality crystals through grid-screening were largely unsuccessful. Varying protein concentration, precipitant, salt, and buffer concentration, and temperature resulted in formation of heavy precipitate in each well. As a result, future efforts towards the crystallization of truncated Smp2 should be directed at additives aimed at increasing protein solubility to allow for formation of crystals over non-native protein aggregates (182).

Table A2.3. Initial truncated Smp2 crystal hits from high-throughput crystallization screens.

Well	Condition	[Smp2] mg/mL	Drop ratio (prot.: well)	Crystal Description
Index HT	0.2 M Calcium chloride dihydrate 0.1 M BIS-TRIS pH 6.5 45% w/v (+/-)-2-Methyl-2,4-pentanediol	15	1:1	microcrystals
PegRX	0.05 M Calcium chloride dihydrate 0.1 M MES monohydrate pH 6.0 45% w/v Polyethylene Glycol 200	15	1:1	microcrystals
MCSG-1	0.16 M Magnesium Chloride 0.08 M Tris: HCl, pH 8.5 24% (w/v) PEG 4000 20% (v/v) glycerol	15	1:1	small crystal
Peg/Ion	0.2 M Magnesium Chloride Hexahydrate 20% w/v PEG 3350	15	1:1	small crystal

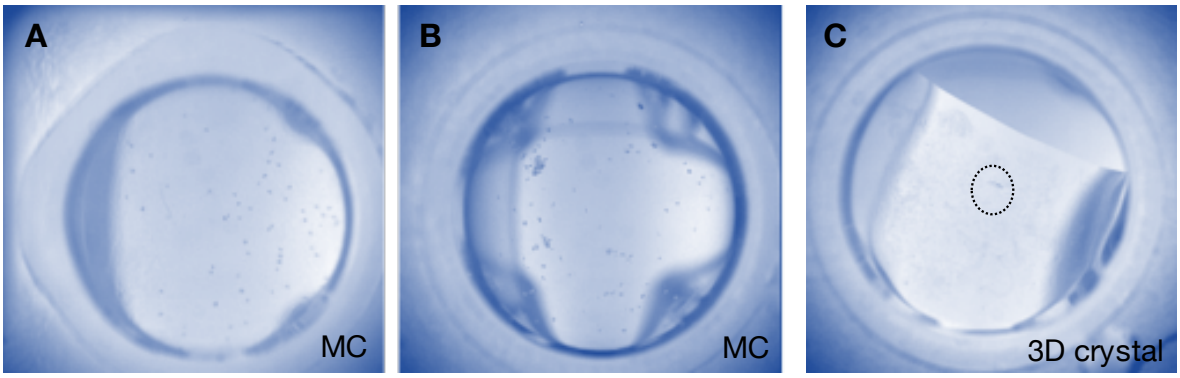

Figure A2. 7. Representative crystal hits of truncated Smp2 from initial HT crystallization condition screening. A, Microcrystals (MC) obtained from 0.2 M $CaCl_2 \cdot 2H_2O$, 0.1 M BisTris pH 6.5, 45% w/v (+/-)-2-Methyl-2,4-pentanediol. **B,** MC obtained from 0.05 M $CaCl_2 \cdot 2H_2O$, 0.1 M MES monohydrate pH 6.0, 45% w/v PEG 200. **C,** Small 3D crystal (circled) obtained from 0.16 M $MgCl_2$, 0.08 M Tris-HCl pH 8.5, 24% w/v PEG 4000, 20% v/v glycerol.

A2.3.3. Low-resolution molecular envelopes for Nir2 and Lipin-1β

In parallel to crystallization condition screening for Nir2, small-angle x-ray scattering (SAXS) studies of Nir2 were performed to obtain a low-resolution molecular envelope of Nir2 in solution. Guinier analysis of Nir2 scattering intensity profiles indicated high-quality data and sufficient linearity at small-angles allowing for determination of the R_g (R_g = 31.3 ± 0.669 Å) and cross-sectional R_g (R_c = 13.8 ± 0.378 Å) for the particle in solution (figure A2.8B). Fourier transformation of the scattering intensity profiles allowed calculation of the distance distribution function for Nir2 at 1.9 mg/mL gave a maximum inter-particle dimension, D_{max}, of 103.49 Å (figure A2.8C). Cylindrical approximations (equation 5.6), the R_g and cross-sectional R_g (R_c) suggest dimensions of radius of 19.65 Å and height of 100.23Å for the Nir2 particle. Notably, the consistency of the maximum dimensions calculated through Guinier analysis and P(r) analyses (~100 Å) provides independent confirmation of data quality and particle dimensions. Guinier and P(r) analyses indicated Nir2 SAXS intensity profiles were of sufficiently high quality for shape-reconstruction. *Ab initio* bead modeling algorithms were applied to the data via DAMMIN (183) to produce an averaged low-resolution molecular envelope for Nir2 (figure A2.8D). Analysis of the molecular dimensions of the real-space shape-reconstruction envelope revealed

dimensions similar, but not identical, to those calculated from the reciprocal-space R_g and cylindrical approximation calculations. Notably, the molecular envelope of Nir2 appears to have 2-fold symmetry indicating the possible formation of a Nir2 dimer in solution.

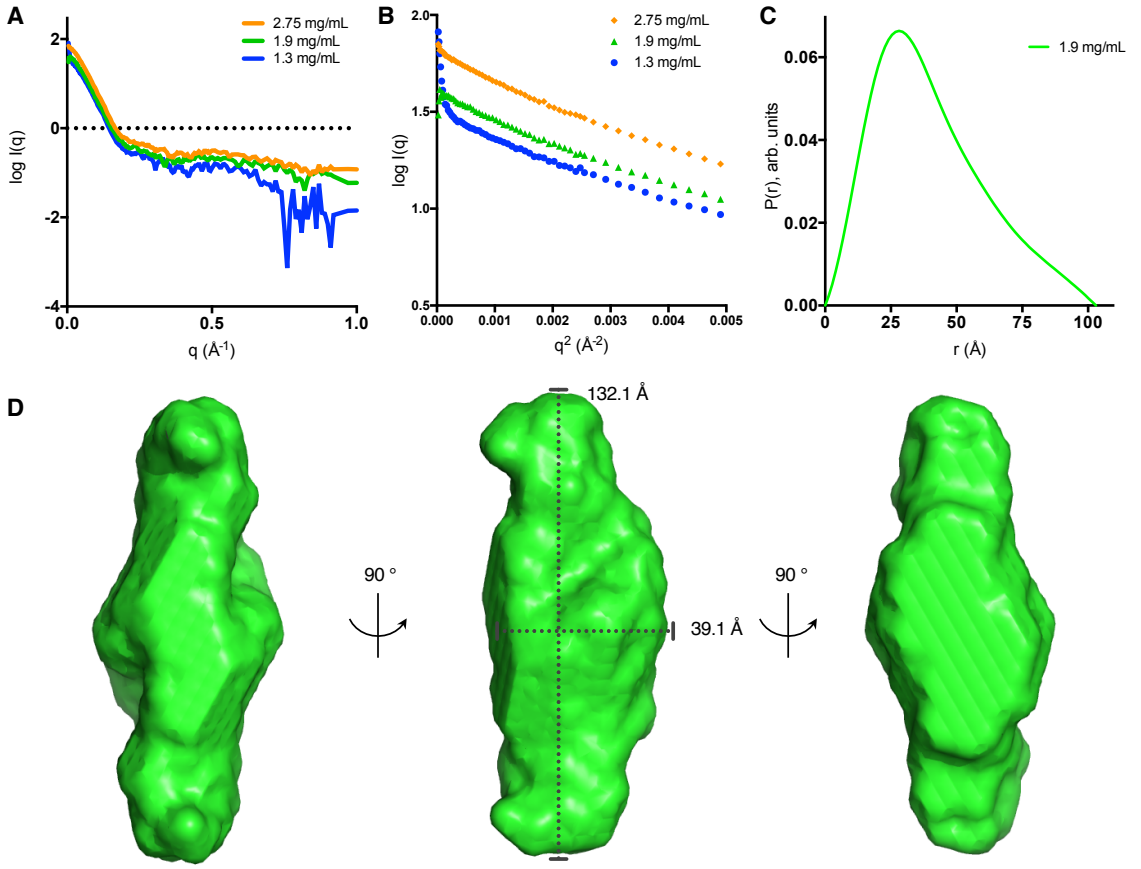

Figure A2.8. SAXS experiments allow calculation of the molecular envelope of Nir2. A, SAXS intensity profiles for Nir2 at 1.3 mg/mL (blue), 1.9 mg/mL (green), and 2.75 mg/mL (orange). **B,** Guinier analyses of Nir2 data at the three concentrations. Data at 2.75 mg/mL (orange) and 1.9 mg/mL (green) are linear within the Guinier region. Aggregation is apparent in the 1.3 mg/mL sample (blue) as the sharp increase in intensity at very small angles. **C,** P(r) function for the 1.9 mg/ mL SAXS dataset indicating a D_{max} of ~103 Å. **D,** *Ab inito* shape-reconstruction of the molecular envelop of Nir2 from the P(r)-transformed data in panel C.

In addition, SAXS was used to investigate the oligomeric state, volume, and shape of the lipin-1β particle prepared from HeLa cells by James Eaton (Harris Lab, University of Virginia). Guinier analysis (126) allowed determination of the R_g to be equal to 64.8 ± 1 Å from the linear region at small-angles (Figure A2.9B). Fourier transformation of the scattering intensity profile resulted in the calculation of the the P(r) function. The x-intercept of the P(r) function corresponds to the maximum diameter for the particle in solution (D_{max} = ~230 Å) (figure A2.9 C). *Ab initio* shape-reconstruction (183) of the lipin-1β particle shows a ring-like structure with an estimated volume of 4.3 x 10^7 $Å^3$ (Figure A2.9D). Notably, the ring structure has a diameter of approximately 60 Å which is similar to the dimension estimated for the center diameter of ring-like particles observed negative-stain EM images collected by the Harris Lab (data not shown, unpublished results). From these results we expect the particle of lipin-1β to be multimeric with at least two subunits.

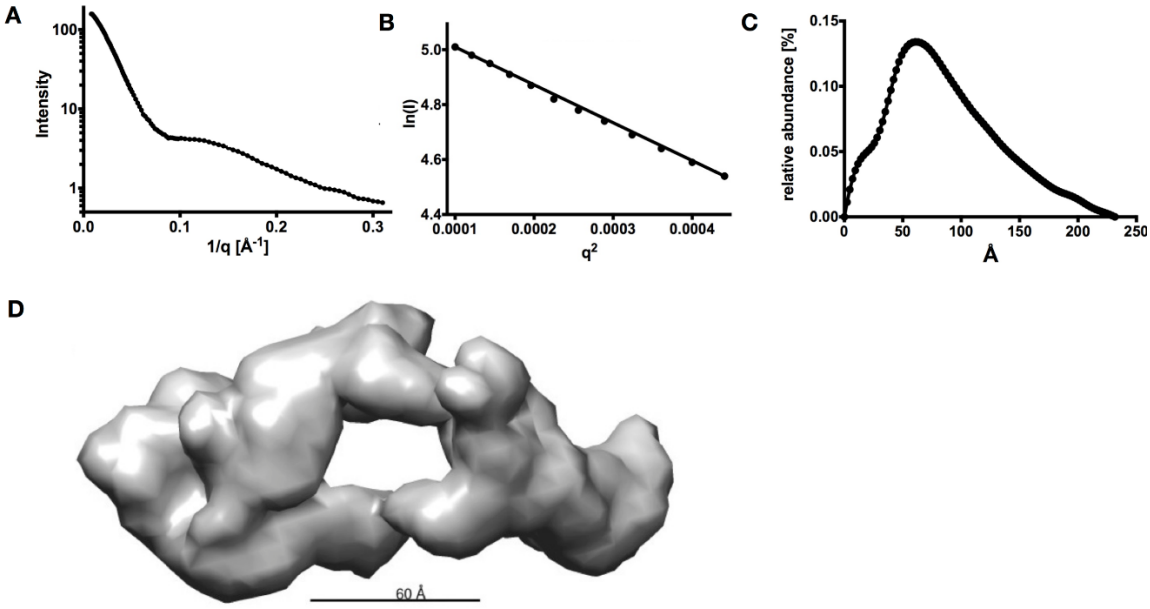

Figure A2.9. SAXS experiments allow calculation of the molecular envelope of human lipin-1β. A, SAXS intensity profile for lipin-1β. **B,** Guinier analyses of lipin-1β data showing linearity within the Guinier region allowing for the calculation of R_g = 64.8 ± 1 Å. **C,** P(r) function shows ring-like topology with maxima at 25 Å and 60 Å and D_{max} of ~230 Å is observed. **D,** *Ab inito* shape-reconstruction of the molecular envelop of Lipin-1β from the P(r) transformed data in panel C shows ring-like topology with an ~60 Å opening at the center. Data shown in this figure processed by Dr. Stefan Jehle (Allen Lab, BU).

A2.3.4. AUC experiments with lipin-1β

In order to validate molecular envelope for lipin-1β determined by SAXS via an orthogonal biophysical method, analytical ultracentrifugation (AUC) sedimentation velocity (SV) and equilibrium (SE) experiments were performed. Fitting the SV data to Lamm equations (174) describing particle sedimentation is

best when using a Monomer-dimer-tetramer model with molecular weights for the species of 72.8 kDa, 217 kDa, and 402 kDa, respectively (figure A2.10 A, B). These molecular weights are lower than expected with the molecular weight of the monomer is calculated to be 102 kDa. The under-estimation of molecular weight suggests the presence of lipids moving with the protein, resulting in an effective lowering of the buoyant molecular mass in solution (184). Overall, the non-standard sedimentation behavior observed for both SV and SE-AUC experiments with lipin-1β suggests lipids are either tightly bound or covalently linked to lipin.

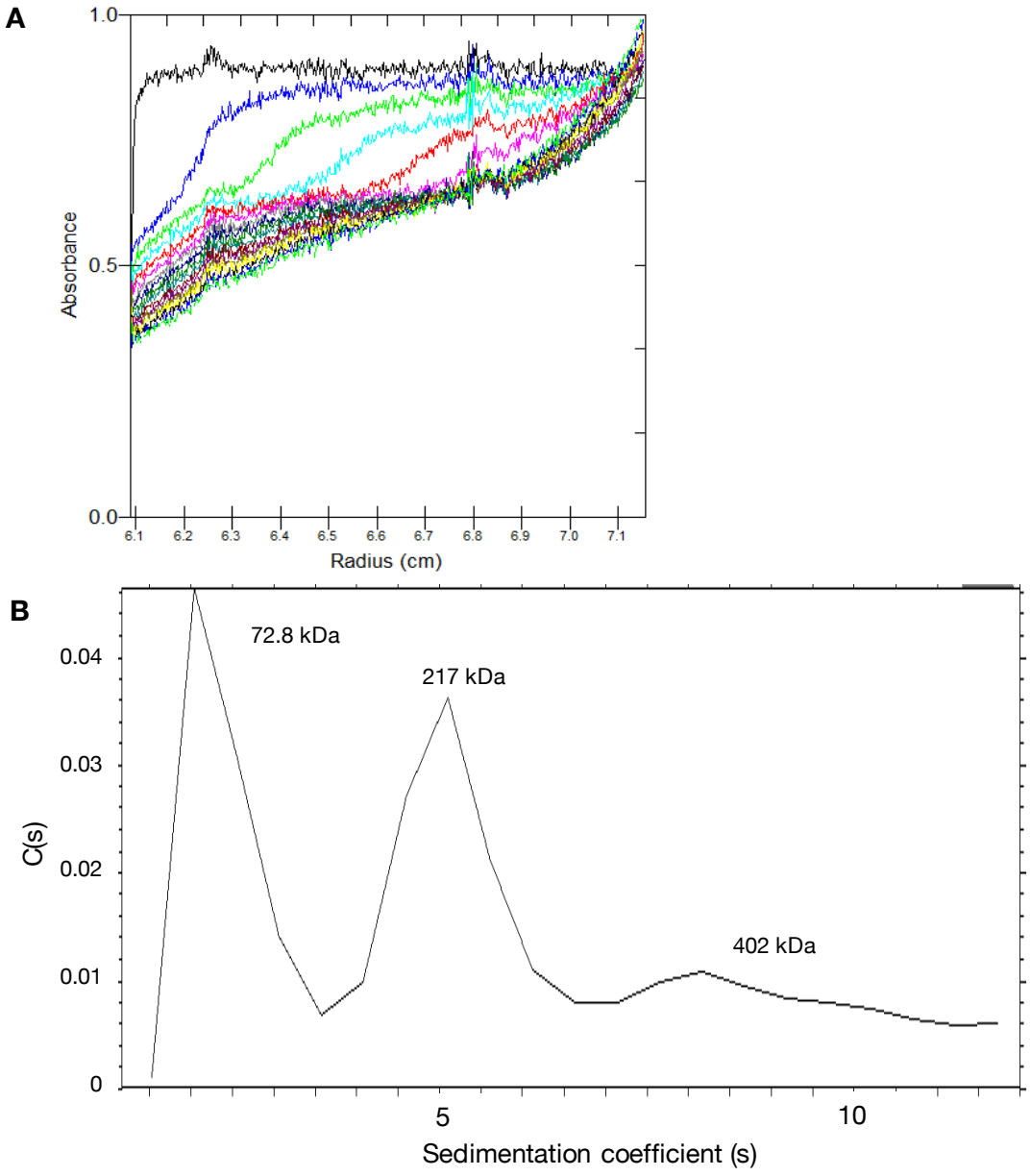

Figure A2.10. SV-AUC experiments with lipin-1β. A, Representative experimental data for lipin-1β in SV-AUC. **B,** Resulting C(s) distribution from fitting the data to a Lamm equation model of self-interacting monomer-dimer-trimer species. Given MW correspond to the estimation of MW for each peak.

To further investigate the presence of lipids tightly bound or covalently-linked to lipin mass spectrometry (MS) experiments were performed in collaboration with the Costello group at Boston University Medical Center. Specifically, liquid chromatography-MS (LC-MS) was attempted to look for lipid modifications to Lipin1. Tryptic digestion of the protein, using sequencing grade trypsin (Promega), was performed on 100 micrograms of sample in 100 mM ammonium bicarbonate in the presence of 1 mM dithiothreitol (DTT), 5 uM iodoacetamide, and 7.5% trifluoroethanol. Trypsin was added in a 1:20 ratio with lipin-1β and digested overnight. Sample was stored at -80 prior to running LC-MS. From this initial experiment, 66% sequence coverage of the peptides identified from the mass spectrum was obtained. However, none of the hypothesized lipid modifications were observed. In the future forward, repletion of the LC-MS experiments with an increased protein amount as well as a more hydrophobic column (C4) could allow for enrichment for possible non-covalently bound lipids. Additionally, analysis of the lipin-1β sequence in collaboration with Dr. Cathy Costello (BUMC) further with the aid of Dr. Cathy Costello, revealed a number of canonical N-linked glycosylation sites (S-X-P) throughout the protein, namely concentrated in the 540aa intervening domain of lipin-1β (185, 186). It is possible that large amounts of glycosylation can affect the AUC data in the way we have

observed (187). It would be worthwhile to investigate the presence of glycopeptides through MS analysis of lipin treated with n-glycosidase and subsequently run through matrix assisted laser desorption and ionization- mass spectrometry (MALDI-MS).

A2.3.5. Microscale thermophoresis investigation of lipin binding interactions

Mircoscale thermophoresis (MST) provides a unique avenue for studying high-affinitity binding interactions of lipin-1 with its binding-partners and the membrane, overall (188). In the preliminary MST experiments, lipin-1β was fluorescently labeled at lysine residues (approximately 5 lysines/ lipin). When the labeled lipin (held at a constant 1.8 μM) was titrated with increasing amounts of unlabeled purified PPARα (0.3 nM-5800 nM) in 100 mM Tris buffer, pH 8.0 the thermophoretic movement markedly changed as indicated by an increase in fluorescence. This change in thermophoretic movement results in a sigmoidal protein-protein binding curve (figure A2.11A). From fitting the resultant data to the Hill equation, the calculation of the equilibrium dissociation constant revealed a K_d of 16±7 nM. This dissociation constant is within the physiological range observed for transcriptional co-regulators binding to PPARα (between 10^{-7}-10^{-10} M) (189). MST was also used to probe for binding between full-length lipin1, which

possesses N-terminal amphipathic helix hypothesized to bind to the membrane, and nanodisc phospholipid membrane bilayers *in vitro*. To this end, full length lipin1β was labeled at lysine residues with a 5x excess of NT-495 dye. This fluorescent lipin-1β was then titrated with increasing amounts of membrane lipid bilayer nanodiscs with a diameter of 8 nm (MSPΔH5, DMPC/DMPG 3:1) (concentration range 0.08 nM- 11 μM) in 100 mM Tris buffer pH 8.0. The resulting sigmoidal binding curve suggests binding between lipin1 and the membrane nanodiscs with a K_d of 21 nM. While this binding constant is dependent upon the number of available binding sites per nanodiscs, there is however, a clear difference between the bound and unbound state of lipin-1β with the nanodiscs (figure A2.11B).

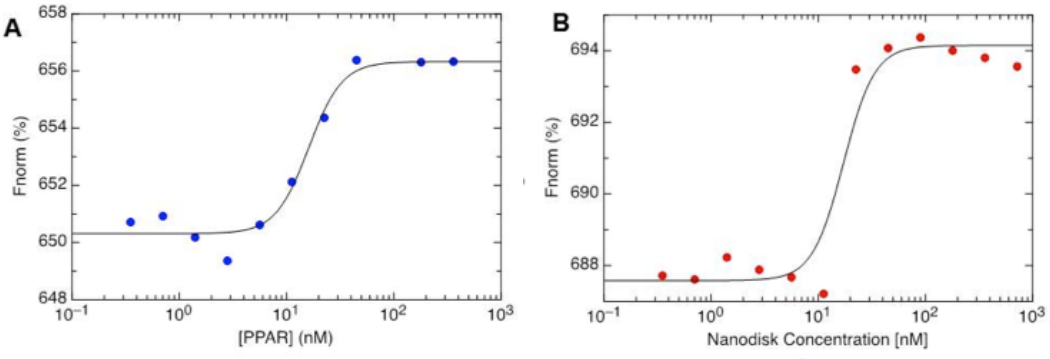

Figure A2.11. MST reveal tight-binding interactions between lipin-1β and PPARα and membrane-bilayer nanodiscs. A, Labeled lipin-1β (held at a constant 1.8 μM) titrated with increasing amounts of unlabeled purified PPARα (0.3 nM-5800 nM) in 100 mM Tris buffer, pH 8.0. Fitting data to the hill equation gave K_d = 16 ± 7 nM. **B,** Labeled lipin-1β was then titrated with increasing amounts of

membrane lipid bilayer nanodiscs (0.08 nM- 11μM) in 100 mM Tris buffer pH 8.0. Fitting data to the hill equation gave $K_d \cong 21$ nM.

A2.4. Conclusions and Future Directions

Overall, optimization of the purification protocols for Nir2 and truncated Smp2 yielded expression and purification of high-quality protein for crystallographic studies of the structural determinants of lipin-1 cellular localization and function. However, recent purifications of Smp2 had increasingly poor yields. Because the truncated Spm2 construct is co-transformed with the pRare plasmid, it will be worthwhile to have the gene synthesized independently to allow for optimal expression in the future. Nir2 expression remains promising, however no crystallization condition screens yielded viable crystals for high-resolution diffraction. Nir2 would be a good candidate for high-throughput crystallization condition screening at the Hauptman-Woodward Institute to explore vast areas of chemical space with minimal amounts of Nir2 protein.

SAXS experiments of Nir2 afforded a molecular envelope which displayed 2-fold symmetry and molecular dimensions that could be consistent with formation of a Nir2 dimer in solution. Similarly, SAXS experiments with human-cell-derived lipin-1β afforded a molecular envelope, which corroborated negative-stain EM molecular dimensions. In an attempt to validate the oligomeric state of

the SAXS envelope, AUC data experiments were performed. However, the sedimentation behavior of lipin in these experiments was highly non-ideal, suggesting the presence of tightly-bound or covalently-bound lipids resulting in lower effective-molecular weights. LC-MS experiments aimed at determining the identity of any bound lipids to lipin were unsuccessful, however analysis of the lipin-1β sequences suggested high prevalence of canonical glycosylation sites in the lipin intervening domain, and across the sequence as a whole.

Preliminary experiments investigating the binding interactions with lipin-1β with its predicted binding-partner PPARα and the lipid-bilayer membrane via experiments with nanodiscs were performed via MST. These initial experiments allowed for determination of preliminary binding constants for these associations. MST presents a unique opportunity to study native interactions of lipin with its binding partners in that it does not require tagged proteins and will allow for easy titration of a complex with a third member of a ternary complex. Therefore, the predicted associations between lipin-1, PPARα, and PGC-1α can be investigated. Utilization of truncation constructs of lipin and PPARα will allow determination of the specific binding sites between the two partners by assessing the impact of removing binding-site motifs by deletion or mutagenesis on the K_d. Similar deletion and mutagenesis studies can be investigated by MST to determine the

mode of membrane association for lipin-1β with the lipid-bilayer nanodiscs. Overall, the experiments detailed in this chapter and proposed for the future will allow for a molecular understanding of the structural determinants of lipin-1 cellular localization and function.

LIST OF JOURNAL ABBREVIATIONS

Acta crysatllogr Sect F Struct Biol Comm .....

.....Acta Crystallographica Section F Structural Biology Communications

Acta Crystallogr D Biol Crystallogr .....

..... Acta Crystallographica Section D Biological Crystallography

Anal Biochem.....Analytical Biochemistry

Anal Chem .....Analytical Chemistry

Angew Chem Int Ed Engl. .....Angewandte Chemie International Edition English

Arch Biochem Biophys..... Archives in Biochemistry and Biophysics

Biochim Biophys Acta ..... Biochimica et Biophysica Acta

Biophys J.....Biophysical Journal

BMC Bioinformatics.....BioMed Central Bioinformatics

BMC Struct Biol ..... BioMed Central Structural Biology

Cell Metab ..... Cell Metabolism

Chem Biol ..... Chemical Biology

Chem Phys Lipids. .....Chemistry and Physics of Lipids

Chem Rev .....Chemical Reviews

Cryst Growth Des .....Crystal Growth and Design

Curr Opin Biotechnol ..... Current Opinions in Biotechnology

Curr Opin Struct Biol..... Current Opinions in Structural Biology

EMBO J ..... European Molecular Biology Organization Journal

EMBO Rep..... European Molecular Biology Organization Reports

Expert Opin Drug Discov .....Expert Opinions in Drug Discovery

FASEB J.....Federation of American Societies for Experimental Biology Journal

FEBS Lett .....Federation of European Biochemical Societies Letters

J Am Chem Soc. .....Journal of the American Chemical Society

J Appl Crystallogr ..... Journal of Applied Crystallography

J Bacteriol.....Journal of Bacteriology

J Biol Chem.....Journal of Biological Chemistry

J Biomol Struct Dyn ..... Journal of Biomolecular Structure and Dynamics

J Comput Chem.....Journal of Computational Chemistry

J Gen Physiol. ..... Journal of General Physiology

J Mol Biol .....Journal of Molecular biology

J Mol Biol ..... Journal of Molecular Biology

J Phys Chem B ..... Journal of Physical Chemistry B

J Struct Biol.....Journal of Structural Biology

J Synhrotron Radiat. .....Journal of Synchrotron Radiation

Methods Enzymol. .....Methods in Enzymology

Microbiol Mol Biol Rev ..... Microbiology and Molecular Biology Reviews

Mol cell proteomics.....Molecular Cell Proteomics

Mol Mircobiol .....Molecular Microbiology

Nat Chem Biol .....Nature Chemical Biology

Nat Commun .....Nature Communications

Nat Genet .....Nature Genetics

Nat Prod Rep .....Natural Product Reports

Nat Protoc. ..... Nature Protocols

Nat Rev Microbiol..... Nature Reviews Microbiology

Nat Rev Mol Cell Biol.....Nature Reviews Molecular and Cell Biology

Nucleic acids res.....Nucleic Acids Research

Plos Comput Biol ..... Public Library of Science Computational Biology

Proc Natl Acad Sci USA ....Proceedings of the National Academy of Sciences USA

Prog Lipid Res ..... Progress in Lipid Research

Protein Eng.....Protein Engineering

Protein Expr Purif .....Protein Expression and Purification

Protein Sci.....Protein Science

Rev Infect Dis..... Reviews of Infectious Diseases

Sci Rep .....Science Reports

Trends Biochem Sci..... Trends in Biochemical Sciences

REFERENCES

1. Spiro RG. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. Glycobiology. 2002;12(4):43R-56R. Epub 2002/06/04. PubMed PMID: 12042244.

2. Jones MB, Rosenberg JN, Betenbaugh MJ, Krag SS. Structure and synthesis of polyisoprenoids used in N-glycosylation across the three domains of life. Biochim Biophys Acta. 2009;1790(6):485-94. Epub 2009/04/08. doi: 10.1016/j.bbagen.2009.03.030. PubMed PMID: 19348869; PMCID: PMC2755495.

3. Varki A. Biological roles of oligosaccharides: all of the theories are correct. Glycobiology. 1993;3(2):97-130. Epub 1993/04/01. PubMed PMID: 8490246.

4. Tytgat HL, Lebeer S. The sweet tooth of bacteria: common themes in bacterial glycoconjugates. Microbiol Mol Biol Rev. 2014;78(3):372-417. Epub 2014/09/04. doi: 10.1128/MMBR.00007-14. PubMed PMID: 25184559; PMCID: PMC4187687.

5. Alemka A, Nothaft H, Zheng J, Szymanski CM. N-glycosylation of Campylobacter jejuni surface proteins promotes bacterial fitness. Infection and immunity. 2013;81(5):1674-82. Epub 2013/03/06. doi: 10.1128/iai.01370-12. PubMed PMID: 23460522; PMCID: PMC3648013.

6. Szymanski CM, Burr DH, Guerry P. Campylobacter protein glycosylation affects host cell interactions. Infection and immunity. 2002;70(4):2242-4. Epub 2002/03/16. PubMed PMID: 11895996; PMCID: PMC127875.

7. Weerapana E, Imperiali B. Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. Glycobiology. 2006;16(6):91R-101R. Epub 2006/03/03. doi: 10.1093/glycob/cwj099. PubMed PMID: 16510493.

8. Jann K, Jann B. Polysaccharide antigens of Escherichia coli. Rev Infect Dis. 1987;9 Suppl 5:S517-26. Epub 1987/09/01. PubMed PMID: 2446369.

9. Karlyshev AV, Everest P, Linton D, Cawthraw S, Newell DG, Wren BW. The Campylobacter jejuni general glycosylation system is important for attachment to human epithelial cells and in the colonization of chicks.

Microbiology. 2004;150(Pt 6):1957-64. Epub 2004/06/09. doi: 10.1099/mic.0.26721-0. PubMed PMID: 15184581.

10. Young NM, Brisson JR, Kelly J, Watson DC, Tessier L, Lanthier PH, Jarrell HC, Cadotte N, St Michael F, Aberg E, Szymanski CM. Structure of the N-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, Campylobacter jejuni. J Biol Chem. 2002;277(45):42530-9. Epub 2002/08/21. doi: 10.1074/jbc.M206114200. PubMed PMID: 12186869.

11. Linton D, Dorrell N, Hitchen PG, Amber S, Karlyshev AV, Morris HR, Dell A, Valvano MA, Aebi M, Wren BW. Functional analysis of the Campylobacter jejuni N-linked protein glycosylation pathway. Mol Microbiol. 2005;55(6):1695-703. Epub 2005/03/09. doi: 10.1111/j.1365-2958.2005.04519.x. PubMed PMID: 15752194.

12. Glover KJ, Weerapana E, Imperiali B. In vitro assembly of the undecaprenylpyrophosphate-linked heptasaccharide for prokaryotic N-linked glycosylation. Proc Natl Acad Sci U S A. 2005;102(40):14255-9. Epub 2005/09/28. doi: 10.1073/pnas.0507311102. PubMed PMID: 16186480; PMCID: PMC1242339.

13. Glover KJ, Weerapana E, Numao S, Imperiali B. Chemoenzymatic synthesis of glycopeptides with PglB, a bacterial oligosaccharyl transferase from Campylobacter jejuni. Chem Biol. 2005;12(12):1311-5. Epub 2005/12/17. doi: 10.1016/j.chembiol.2005.10.004. PubMed PMID: 16356848; PMCID: PMC2519243.

14. Price NP, Momany FA. Modeling bacterial UDP-HexNAc: polyprenol-P HexNAc-1-P transferases. Glycobiology. 2005;15(9):29R-42R. Epub 2005/04/22. doi: 10.1093/glycob/cwi065. PubMed PMID: 15843595.

15. Lukose V, Walvoort MTC, Imperiali B. Bacterial phosphoglycosyl transferases: initiators of glycan biosynthesis at the membrane interface. Glycobiology. 2017;27(9):820-33. Epub 2017/08/16. doi: 10.1093/glycob/cwx064. PubMed PMID: 28810664.

16. Hartley MD, Imperiali B. At the membrane frontier: a prospectus on the remarkable evolutionary conservation of polyprenols and polyprenyl-phosphates. Arch Biochem Biophys. 2012;517(2):83-97. Epub 2011/11/19. doi: 10.1016/j.abb.2011.10.018. PubMed PMID: 22093697; PMCID: PMC3253937.

17. Lukose V, Luo L, Kozakov D, Vajda S, Allen KN, Imperiali B. Conservation and Covariance in Small Bacterial Phosphoglycosyltransferases Identify the Functional Catalytic Core. Biochemistry. 2015;54(50):7326-34. doi: 10.1021/acs.biochem.5b01086.

18. Lehrer J, Vigeant KA, Tatar LD, Valvano MA. Functional characterization and membrane topology of Escherichia coli WecA, a sugar-phosphate transferase initiating the biosynthesis of enterobacterial common antigen and O-antigen lipopolysaccharide. J Bacteriol. 2007;189(7):2618-28. Epub 2007/01/24. doi: 10.1128/JB.01905-06. PubMed PMID: 17237164; PMCID: PMC1855806.

19. Chung BC, Zhao J, Gillespie RA, Kwon DY, Guan Z, Hong J, Zhou P, Lee SY. Crystal structure of MraY, an essential membrane enzyme for bacterial cell wall synthesis. Science. 2013;341(6149):1012-6. Epub 2013/08/31. doi: 10.1126/science.1236501. PubMed PMID: 23990562; PMCID: PMC3906829.

20. Glover KJ, Weerapana E, Chen MM, Imperiali B. Direct biochemical evidence for the utilization of UDP-bacillosamine by PglC, an essential glycosyl-1-phosphate transferase in the Campylobacter jejuni N-linked glycosylation pathway. Biochemistry. 2006;45(16):5343-50. Epub 2006/04/19. doi: 10.1021/bi0602056. PubMed PMID: 16618123.

21. Hartley MD, Morrison MJ, Aas FE, Borud B, Koomey M, Imperiali B. Biochemical characterization of the O-linked glycosylation pathway in Neisseria gonorrhoeae responsible for biosynthesis of protein glycans containing N,N'-diacetylbacillosamine. Biochemistry. 2011;50(22):4936-48. Epub 2011/05/06. doi: 10.1021/bi2003372. PubMed PMID: 21542610; PMCID: PMC3108506.

22. Saldias MS, Patel K, Marolda CL, Bittner M, Contreras I, Valvano MA. Distinct functional domains of the Salmonella enterica WbaP transferase that is involved in the initiation reaction for synthesis of the O antigen subunit. Microbiology. 2008;154(Pt 2):440-53. Epub 2008/01/30. doi: 10.1099/mic.0.2007/013136-0. PubMed PMID: 18227248.

23. Al-Dabbagh B, Mengin-Lecreulx D, Bouhss A. Purification and characterization of the bacterial UDP-GlcNAc:undecaprenyl-phosphate GlcNAc-1-phosphate transferase WecA. J Bacteriol. 2008;190(21):7141-6. Epub 2008/08/30. doi: 10.1128/JB.00676-08. PubMed PMID: 18723618; PMCID: PMC2580700.

24. Al-Dabbagh B, Olatunji S, Crouvoisier M, El Ghachi M, Blanot D, Mengin-Lecreulx D, Bouhss A. Catalytic mechanism of MraY and WecA, two paralogues of the polyprenyl-phosphate N-acetylhexosamine 1-phosphate transferase superfamily. Biochimie. 2016;127:249-57. Epub 2016/06/18. doi: 10.1016/j.biochi.2016.06.005. PubMed PMID: 27312048.

25. Das D, Kuzmic P, Imperiali B. Analysis of a dual domain phosphoglycosyl transferase reveals a ping-pong mechanism with a covalent enzyme intermediate. Proc Natl Acad Sci U S A. 2017;114(27):7019-24. Epub 2017/06/21. doi: 10.1073/pnas.1703397114. PubMed PMID: 28630348; PMCID: PMC5502628.

26. Winn M, Goss RJ, Kimura K, Bugg TD. Antimicrobial nucleoside antibiotics targeting cell wall assembly: recent advances in structure-function studies and nucleoside biosynthesis. Nat Prod Rep. 2010;27(2):279-304. Epub 2010/01/30. doi: 10.1039/b816215h. PubMed PMID: 20111805.

27. Keller RK, Boon DY, Crum FC. N-Acetylglucosamine- 1 -phosphate transferase from hen oviduct: solubilization, characterization, and inhibition by tunicamycin. Biochemistry. 1979;18(18):3946-52. Epub 1979/09/04. PubMed PMID: 486403.

28. Morrison MJ, Imperiali B. Biochemical analysis and structure determination of bacterial acetyltransferases responsible for the biosynthesis of UDP-N,N'-diacetylbacillosamine. J Biol Chem. 2013;288(45):32248-60. Epub 2013/09/26. doi: 10.1074/jbc.M113.510560. PubMed PMID: 24064219; PMCID: PMC3820863.

29. Walvoort MT, Lukose V, Imperiali B. A Modular Approach to Phosphoglycosyltransferase Inhibitors Inspired by Nucleoside Antibiotics. Chemistry. 2016;22(11):3856-64. Epub 2015/12/15. doi: 10.1002/chem.201503986. PubMed PMID: 26662170; PMCID: PMC5506376.

30. Liu J, Rost B. Comparing function and structure between entire proteomes. Protein Sci. 2001;10(10):1970-9. Epub 2001/09/22. doi: 10.1110/ps.10101. PubMed PMID: 11567088; PMCID: PMC2374214.

31. Wagner S, Baars L, Ytterberg AJ, Klussmeier A, Wagner CS, Nord O, Nygren PA, van Wijk KJ, de Gier JW. Consequences of membrane protein

overexpression in Escherichia coli. Mol Cell Proteomics. 2007;6(9):1527-50. Epub 2007/04/21. doi: 10.1074/mcp.M600431-MCP200. PubMed PMID: 17446557.

32. Wiener MC. A pedestrian guide to membrane protein crystallization. Methods. 2004;34(3):364-72. Epub 2004/08/25. doi: 10.1016/j.ymeth.2004.03.025. PubMed PMID: 15325654.

33. Carpenter EP, Beis K, Cameron AD, Iwata S. Overcoming the challenges of membrane protein crystallography. Curr Opin Struct Biol. 2008;18(5):581-6. Epub 2008/08/05. doi: 10.1016/j.sbi.2008.07.001. PubMed PMID: 18674618; PMCID: PMC2580798.

34. Lukose V. Priming and processing glycosyltransferases in bacterial N-linked glycosylation pathways [Doctoral Dissertation]. Cambridge, MA: Massachusetts Institute of Technology; 2015. http://hdl.handle.net/1721.1/101550

35. Lynch AJ. Ligand selectivity: binding at the protein-protein interface of KEAP1 and NEMO [Doctoral Dissertation]. Boston, MA: Boston University; 2015. https://search-proquest-com.ezproxy.bu.edu/docview/1767228336?accountid=9676

36. Taylor GL. Introduction to phasing. Acta Crystallogr D Biol Crystallogr. 2010;66(Pt 4):325-38. Epub 2010/04/13. doi: 10.1107/S0907444910006694. PubMed PMID: 20382985; PMCID: PMC2852296.

37. Rossmann MG, Blow DM. The detection of sub-units within the crystallographic asymmetric unit. Acta Crystallogr D Biol Crystallogr. 1962;15:24-31.

38. McCoy AJ. Solving structures of protein complexes by molecular replacement with Phaser. Acta Crystallogr D Biol Crystallogr. 2007;63(Pt 1):32-41. Epub 2006/12/14. doi: 10.1107/S0907444906045975. PubMed PMID: 17164524; PMCID: PMC2483468.

39. Wang BC. Resolution of phase ambiguity in macromolecular crystallography. Methods Enzymol. 1985;115:90-112. Epub 1985/01/01. PubMed PMID: 4079800.

40. Pike AC, Garman EF, Krojer T, von Delft F, Carpenter EP. An overview of heavy-atom derivatization of protein crystals. Acta Crystallogr D Struct Biol. 2016;72(Pt 3):303-18. Epub 2016/03/10. doi: 10.1107/S2059798316000401. PubMed PMID: 26960118; PMCID: PMC4784662.

41. Doutch J, Hough MA, Hasnain SS, Strange RW. Challenges of sulfur SAD phasing as a routine method in macromolecular crystallography. J Synchrotron Radiat. 2012;19(Pt 1):19-29. Epub 2011/12/22. doi: 10.1107/S0909049511049004. PubMed PMID: 22186640.

42. Hendrickson WA, Horton JR, LeMaster DM. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. EMBO J. 1990;9(5):1665-72. Epub 1990/05/01. PubMed PMID: 2184035; PMCID: PMC551863.

43. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr. 2010;66(Pt 2):213-21. Epub 2010/02/04. doi: 10.1107/S0907444909052925. PubMed PMID: 20124702; PMCID: PMC2815670.

44. Kabsch W. Xds. Acta Crystallogr D Biol Crystallogr. 2010;66(Pt 2):125-32. Epub 2010/02/04. doi: 10.1107/S0907444909047337. PubMed PMID: 20124692; PMCID: PMC2815665.

45. Schneider TR, Sheldrick GM. Substructure solution with SHELXD. Acta Crystallogr D Biol Crystallogr. 2002;58(Pt 10 Pt 2):1772-9. Epub 2002/09/28. PubMed PMID: 12351820.

46. Terwilliger TC, Berendzen J. Automated MAD and MIR structure solution. Acta Crystallogr D Biol Crystallogr. 1999;55(Pt 4):849-61. Epub 1999/03/25. PubMed PMID: 10089316; PMCID: PMC2746121.

47. Terwilliger TC. Maximum-likelihood density modification. Acta Crystallogr D Biol Crystallogr. 2000;56(Pt 8):965-72. Epub 2000/08/16. PubMed PMID: 10944333; PMCID: PMC2792768.

48. Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Moriarty NW, Zwart PH, Hung LW, Read RJ, Adams PD. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. Acta Crystallogr D Biol Crystallogr. 2008;64(Pt 1):61-9. Epub 2007/12/21. doi: 10.1107/S090744490705024X. PubMed PMID: 18094468; PMCID: PMC2394820.

49. Wang S, Sun S, Li Z, Zhang R, Xu J. Accurate De Novo Prediction of Protein Contact Map by Ultra-Deep Learning Model. PLoS Comput Biol. 2017;13(1):e1005324. Epub 2017/01/06. doi: 10.1371/journal.pcbi.1005324. PubMed PMID: 28056090; PMCID: PMC5249242.

50. Ovchinnikov S, Park H, Varghese N, Huang PS, Pavlopoulos GA, Kim DE, Kamisetty H, Kyrpides NC, Baker D. Protein structure determination using metagenome sequence data. Science. 2017;355(6322):294-8. Epub 2017/01/21. doi: 10.1126/science.aah4043. PubMed PMID: 28104891; PMCID: PMC5493203.

51. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr. 2007;40(Pt 4):658-74. Epub 2007/08/01. doi: 10.1107/S0021889807021206. PubMed PMID: 19461840; PMCID: PMC2483472.

52. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr. 2010;66(Pt 4):486-501. Epub 2010/04/13. doi: 10.1107/S0907444910007493. PubMed PMID: 20383002; PMCID: PMC2852313.

53. Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH, Adams PD. Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr D Biol Crystallogr. 2012;68(Pt 4):352-67. Epub 2012/04/17. doi: 10.1107/S0907444912001308. PubMed PMID: 22505256; PMCID: PMC3322595.

54. Zwart PH. Anomalous signal indicators in protein crystallography. Acta Crystallogr D Biol Crystallogr. 2005;61(Pt 11):1437-48. Epub 2005/10/22. doi: 10.1107/S0907444905023589. PubMed PMID: 16239720.

55. Pape T, Schneider TR. HKL2MAP: a graphical user interface for phasing with SHELX programs. J Appl Cryst. 2004;37:843-4.

56. Sheldrick GM. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. Acta Crystallogr D Biol Crystallogr. 2010;66(Pt 4):479-85. Epub 2010/04/13. doi: 10.1107/S0907444909038360. PubMed PMID: 20383001; PMCID: PMC2852312.

57. Hartley MD, Schneggenburger PE, Imperiali B. Lipid bilayer nanodisc platform for investigating polyprenol-dependent enzyme interactions and activities. Proc Natl Acad Sci U S A. 2013;110(52):20863-70. Epub 2013/12/05. doi: 10.1073/pnas.1320852110. PubMed PMID: 24302767; PMCID: 3876266.

58. Holm L, Rosenstrom P. Dali server: conservation mapping in 3D. Nucleic Acids Res. 2010;38(Web Server issue):W545-9. Epub 2010/05/12. doi: 10.1093/nar/gkq366. PubMed PMID: 20457744; PMCID: PMC2896194.

59. Wallin E, von Heijne G. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. Protein Sci. 1998;7(4):1029-38. Epub 1998/05/06. doi: 10.1002/pro.5560070420. PubMed PMID: 9568909; PMCID: PMC2143985.

60. Krebs MP, Noorwez SM, Malhotra R, Kaushal S. Quality control of integral membrane proteins. Trends Biochem Sci. 2004;29(12):648-55. Epub 2004/11/17. doi: 10.1016/j.tibs.2004.10.009. PubMed PMID: 15544951.

61. Bracey MH, Cravatt BF, Stevens RC. Structural commonalities among integral membrane enzymes. FEBS Lett. 2004;567(2-3):159-65. Epub 2004/06/05. doi: 10.1016/j.febslet.2004.04.084. PubMed PMID: 15178315.

62. Blobel G. Intracellular protein topogenesis. Proc Natl Acad Sci U S A. 1980;77(3):1496-500. Epub 1980/03/01. PubMed PMID: 6929499; PMCID: PMC348522.

63. Nam HJ, Kim I, Bowie JU, Kim S. Metazoans evolved by taking domains from soluble proteins to expand intercellular communication network. Sci Rep. 2015;5:9576. Epub 2015/04/30. doi: 10.1038/srep09576. PubMed PMID: 25923201; PMCID: PMC4894438.

64. Attwood MM, Krishnan A, Almen MS, Schioth HB. Highly diversified expansions shaped the evolution of membrane bound proteins in metazoans. Sci

Rep. 2017;7(1):12387. Epub 2017/09/30. doi: 10.1038/s41598-017-11543-z. PubMed PMID: 28959054; PMCID: PMC5620054.

65. Lomize AL, Pogozheva ID, Lomize MA, Mosberg HI. The role of hydrophobic interactions in positioning of peripheral proteins in membranes. BMC Struct Biol. 2007;7:44. Epub 2007/07/03. doi: 10.1186/1472-6807-7-44. PubMed PMID: 17603894; PMCID: PMC1934363.

66. von Heijne G. Membrane-protein topology. Nat Rev Mol Cell Biol. 2006;7(12):909-18. Epub 2006/12/02. doi: 10.1038/nrm2063. PubMed PMID: 17139331.

67. Balali-Mood K, Bond PJ, Sansom MS. Interaction of monotopic membrane enzymes with a lipid bilayer: a coarse-grained MD simulation study. Biochemistry. 2009;48(10):2135-45. Epub 2009/01/24. doi: 10.1021/bi8017398. PubMed PMID: 19161285.

68. Granseth E, von Heijne G, Elofsson A. A study of the membrane-water interface region of membrane proteins. J Mol Biol. 2005;346(1):377-85. Epub 2005/01/25. doi: 10.1016/j.jmb.2004.11.036. PubMed PMID: 15663952.

69. Speers AE, Wu CC. Proteomics of integral membrane proteins--theory and application. Chem Rev. 2007;107(8):3687-714. Epub 2007/08/09. doi: 10.1021/cr068286z. PubMed PMID: 17683161.

70. Eisenberg D, Schwarz E, Komaromy M, Wall R. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. J Mol Biol. 1984;179(1):125-42. Epub 1984/10/15. PubMed PMID: 6502707.

71. Saier MH, Jr., Tran CV, Barabote RD. TCDB: the Transporter Classification Database for membrane transport protein analyses and information. Nucleic Acids Res. 2006;34(Database issue):D181-6. Epub 2005/12/31. doi: 10.1093/nar/gkj001. PubMed PMID: 16381841; PMCID: PMC1334385.

72. Lomize MA, Pogozheva ID, Joo H, Mosberg HI, Lomize AL. OPM database and PPM web server: resources for positioning of proteins in membranes. Nucleic Acids Res. 2012;40(Database issue):D370-6. Epub 2011/09/06. doi: 10.1093/nar/gkr703. PubMed PMID: 21890895; PMCID: PMC3245162.

73. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci U S A. 2001;98(18):10037-41. Epub 2001/08/23. doi: 10.1073/pnas.181342398. PubMed PMID: 11517324; PMCID: PMC56910.

74. Kumar P, Bansal M. HELANAL-Plus: a web server for analysis of helix geometry in protein structures. J Biomol Struct Dyn. 2012;30(6):773-83. Epub 2012/06/28. doi: 10.1080/07391102.2012.689705. PubMed PMID: 22734588.

75. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem. 2004;25(13):1605-12. Epub 2004/07/21. doi: 10.1002/jcc.20084. PubMed PMID: 15264254.

76. Picot D, Loll PJ, Garavito RM. The X-ray crystal structure of the membrane protein prostaglandin H2 synthase-1. Nature. 1994;367(6460):243-9. Epub 1994/01/20. doi: 10.1038/367243a0. PubMed PMID: 8121489.

77. Bracey MH, Hanson MA, Masuda KR, Stevens RC, Cravatt BF. Structural adaptations in a membrane enzyme that terminates endocannabinoid signaling. Science. 2002;298(5599):1793-6. Epub 2002/12/03. doi: 10.1126/science.1076535. PubMed PMID: 12459591.

78. Wendt KU, Poralla K, Schulz GE. Structure and function of a squalene cyclase. Science. 1997;277(5333):1811-5. Epub 1997/09/20. PubMed PMID: 9295270.

79. Furlong SE, Ford A, Albarnez-Rodriguez L, Valvano MA. Topological analysis of the Escherichia coli WcaJ protein reveals a new conserved configuration for the polyisoprenyl-phosphate hexose-1-phosphate transferase family. Sci Rep. 2015;5:9178. Epub 2015/03/18. doi: 10.1038/srep09178. PubMed PMID: 25776537; PMCID: PMC4361858.

80. Ovchinnikov S, Kamisetty H, Baker D. Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. Elife. 2014;3:e02030. Epub 2014/05/21. doi: 10.7554/eLife.02030. PubMed PMID: 24842992; PMCID: PMC4034769.

81. Nasie I, Steiner-Mordoch S, Schuldiner S. Topology determination of untagged membrane proteins. Methods in molecular biology (Clifton, NJ).

2013;1033:121-30. Epub 2013/09/03. doi: 10.1007/978-1-62703-487-6_8. PubMed PMID: 23996174.

82. Bogdanov M, Zhang W, Xie J, Dowhan W. Transmembrane protein topology mapping by the substituted cysteine accessibility method (SCAM(TM)): application to lipid-specific membrane protein topogenesis. Methods. 2005;36(2):148-71. Epub 2005/05/17. doi: 10.1016/j.ymeth.2004.11.002. PubMed PMID: 15894490; PMCID: PMC4104023.

83. Nilsson I, von Heijne G. Fine-tuning the topology of a polytopic membrane protein: role of positively and negatively charged amino acids. Cell. 1990;62(6):1135-41. Epub 1990/09/21. PubMed PMID: 2119256.

84. Palsdottir H, Hunte C. Lipids in membrane protein structures. Biochim Biophys Acta. 2004;1666(1-2):2-18. Epub 2004/11/03. doi: 10.1016/j.bbamem.2004.06.012. PubMed PMID: 15519305.

85. Gallivan JP, Dougherty DA. Cation-pi interactions in structural biology. Proc Natl Acad Sci U S A. 1999;96(17):9459-64. Epub 1999/08/18. PubMed PMID: 10449714; PMCID: PMC22230.

86. Shirota M, Ishida T, Kinoshita K. Effects of surface-to-volume ratio of proteins on hydrophilic residues: decrease in occurrence and increase in buried fraction. Protein Sci. 2008;17(9):1596-602. Epub 2008/06/17. doi: 10.1110/ps.035592.108. PubMed PMID: 18556475; PMCID: PMC2525519.

87. Janin J. Surface and inside volumes in globular proteins. Nature. 1979;277(5696):491-2. Epub 1979/02/08. PubMed PMID: 763335.

88. Kleywegt GJ, Jones TA. Detection, delineation, measurement and display of cavities in macromolecular structures. Acta Crystallogr D Biol Crystallogr. 1994;50(Pt 2):178-85. Epub 1994/03/01. doi: 10.1107/S0907444993011333. PubMed PMID: 15299456.

89. Pace CN, Scholtz JM. A helix propensity scale based on experimental studies of peptides and proteins. Biophys J. 1998;75(1):422-7. Epub 1998/07/02. PubMed PMID: 9649402; PMCID: PMC1299714.

90. Geetha V. Distortions in protein helices. International Journal of Biological Macromolecules. 1996;19:81-9.

91. Barlow DJ, Thornton JM. Helix geometry in proteins. J Mol Biol. 1988;201(3):601-19. Epub 1988/06/05. PubMed PMID: 3418712.

92. Vieira-Pires RS, Morais-Cabral JH. 3(10) helices in channels and other membrane proteins. J Gen Physiol. 2010;136(6):585-92. Epub 2010/12/01. doi: 10.1085/jgp.201010508. PubMed PMID: 21115694; PMCID: PMC2995148.

93. Pal L, Basu G. Novel protein structural motifs containing two-turn and longer 3(10)-helices. Protein Eng. 1999;12(10):811-4. Epub 1999/11/11. PubMed PMID: 10556239.

94. Karpen ME, de Haseth PL, Neet KE. Differences in the amino acid distributions of 3(10)-helices and alpha-helices. Protein Sci. 1992;1(10):1333-42. Epub 1992/10/01. doi: 10.1002/pro.5560011013. PubMed PMID: 1303752; PMCID: PMC2142095.

95. Allen KN, Dunaway-Mariano D. Catalytic scaffolds for phosphoryl group transfer. Curr Opin Struct Biol. 2016;41:172-9. Epub 2016/08/16. doi: 10.1016/j.sbi.2016.07.017. PubMed PMID: 27526404; PMCID: PMC5154885.

96. Lloyd AJ, Brandish PE, Gilbey AM, Bugg TD. Phospho-N-acetyl-muramyl-pentapeptide translocase from Escherichia coli: catalytic role of conserved aspartic acid residues. J Bacteriol. 2004;186(6):1747-57. Epub 2004/03/05. PubMed PMID: 14996806; PMCID: PMC355978.

97. Szymanski CM, Wren BW. Protein glycosylation in bacterial mucosal pathogens. Nat Rev Microbiol. 2005;3(3):225-37. Epub 2005/03/02. doi: 10.1038/nrmicro1100. PubMed PMID: 15738950.

98. Hakulinen JK, Hering J, Branden G, Chen H, Snijder A, Ek M, Johansson P. MraY-antibiotic complex reveals details of tunicamycin mode of action. Nat Chem Biol. 2017;13(3):265-7. Epub 2017/01/10. doi: 10.1038/nchembio.2270. PubMed PMID: 28068312.

99. Bolam DN, Roberts S, Proctor MR, Turkenburg JP, Dodson EJ, Martinez-Fleites C, Yang M, Davis BG, Davies GJ, Gilbert HJ. The crystal structure of two

macrolide glycosyltransferases provides a blueprint for host cell antibiotic immunity. Proc Natl Acad Sci U S A. 2007;104(13):5336-41. Epub 2007/03/23. doi: 10.1073/pnas.0607897104. PubMed PMID: 17376874; PMCID: PMC1838483.

100. Chang A, Singh S, Phillips GN, Jr., Thorson JS. Glycosyltransferase structural biology and its role in the design of catalysts for glycosylation. Curr Opin Biotechnol. 2011;22(6):800-8. Epub 2011/05/20. doi: 10.1016/j.copbio.2011.04.013. PubMed PMID: 21592771; PMCID: PMC3163058.

101. Lovering AL, Lin LY, Sewell EW, Spreter T, Brown ED, Strynadka NC. Structure of the bacterial teichoic acid polymerase TagF provides insights into membrane association and catalysis. Nat Struct Mol Biol. 2010;17(5):582-9. Epub 2010/04/20. doi: 10.1038/nsmb.1819. PubMed PMID: 20400947.

102. Breukink E, van Heusden HE, Vollmerhaus PJ, Swiezewska E, Brunner L, Walker S, Heck AJ, de Kruijff B. Lipid II is an intrinsic component of the pore induced by nisin in bacterial membranes. J Biol Chem. 2003;278(22):19898-903. Epub 2003/03/29. doi: 10.1074/jbc.M301463200. PubMed PMID: 12663672.

103. Chung BC, Mashalidis EH, Tanino T, Kim M, Matsuda A, Hong J, Ichikawa S, Lee SY. Structural insights into inhibition of lipid I production in bacterial cell wall synthesis. Nature. 2016;533(7604):557-60. Epub 2016/04/19. doi: 10.1038/nature17636. PubMed PMID: 27088606; PMCID: PMC4882255.

104. Chen MM, Weerapana E, Ciepichal E, Stupak J, Reid CW, Swiezewska E, Imperiali B. Polyisoprenol specificity in the Campylobacter jejuni N-linked glycosylation pathway. Biochemistry. 2007;46(50):14342-8. Epub 2007/11/24. doi: 10.1021/bi701956x. PubMed PMID: 18034500; PMCID: PMC2585822.

105. Patel KB, Furlong SE, Valvano MA. Functional analysis of the C-terminal domain of the WbaP protein that mediates initiation of O antigen synthesis in Salmonella enterica. Glycobiology. 2010;20(11):1389-401. Epub 2010/07/02. doi: 10.1093/glycob/cwq104. PubMed PMID: 20591829.

106. Moriarty NW, Grosse-Kunstleve RW, Adams PD. electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. Acta Crystallogr D Biol Crystallogr. 2009;65(Pt 10):1074-80. Epub 2009/09/23. doi: 10.1107/S0907444909029436. PubMed PMID: 19770504; PMCID: PMC2748967.

107. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem. 2009;30(16):2785-91. Epub 2009/04/29. doi: 10.1002/jcc.21256. PubMed PMID: 19399780; PMCID: PMC2760638.

108. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010;31(2):455-61. Epub 2009/06/06. doi: 10.1002/jcc.21334. PubMed PMID: 19499576; PMCID: PMC3041641.

109. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 1997;276:307-26. Epub 1997/01/01. PubMed PMID: 27754618.

110. Amer AO, Valvano MA. Conserved aspartic acids are essential for the enzymic activity of the WecA protein initiating the biosynthesis of O-specific lipopolysaccharide and enterobacterial common antigen in Escherichia coli. Microbiology. 2002;148(Pt 2):571-82. Epub 2002/02/08. doi: 10.1099/00221287-148-2-571. PubMed PMID: 11832520.

111. Schmidt H, Hansen G, Singh S, Hanuszkiewicz A, Lindner B, Fukase K, Woodard RW, Holst O, Hilgenfeld R, Mamat U, Mesters JR. Structural and mechanistic analysis of the membrane-embedded glycosyltransferase WaaA required for lipopolysaccharide synthesis. Proc Natl Acad Sci U S A. 2012;109(16):6253-8. Epub 2012/04/05. doi: 10.1073/pnas.1119894109. PubMed PMID: 22474366; PMCID: PMC3341020.

112. Jervis AJ, Langdon R, Hitchen P, Lawson AJ, Wood A, Fothergill JL, Morris HR, Dell A, Wren B, Linton D. Characterization of N-linked protein glycosylation in Helicobacter pullorum. J Bacteriol. 2010;192(19):5228-36. Epub 2010/06/29. doi: 10.1128/JB.00211-10. PubMed PMID: 20581208; PMCID: PMC2944503.

113. Ardiccioni C, Clarke OB, Tomasek D, Issa HA, von Alpen DC, Pond HL, Banerjee S, Rajashankar KR, Liu Q, Guan Z, Li C, Kloss B, Bruni R, Kloppmann E, Rost B, Manzini MC, Shapiro L, Mancia F. Structure of the polyisoprenyl-phosphate glycosyltransferase GtrB and insights into the mechanism of catalysis.

Nat Commun. 2016;7:10175. Epub 2016/01/06. doi: 10.1038/ncomms10175. PubMed PMID: 26729507; PMCID: PMC4728340.

114. Mattevi A. A monotopic membrane protein goes solo. Structure. 2006;14(4):628-9. Epub 2006/04/18. doi: 10.1016/j.str.2006.03.002. PubMed PMID: 16615901.

115. Rufer AC, Thoma R, Benz J, Stihle M, Gsell B, De Roo E, Banner DW, Mueller F, Chomienne O, Hennig M. The crystal structure of carnitine palmitoyltransferase 2 and implications for diabetes treatment. Structure. 2006;14(4):713-23. Epub 2006/04/18. doi: 10.1016/j.str.2006.01.008. PubMed PMID: 16615913.

116. Luft JR, Snell EH, Detitta GT. Lessons from high-throughput protein crystallization screening: 10 years of practical experience. Expert Opin Drug Discov. 2011;6(5):465-80. Epub 2012/06/01. doi: 10.1517/17460441.2011.566857. PubMed PMID: 22646073; PMCID: PMC3637395.

117. Bijelic A, Molitor C, Mauracher SG, Al-Oweini R, Kortz U, Rompel A. Hen egg-white lysozyme crystallisation: protein stacking and structure stability enhanced by a Tellurium(VI)-centred polyoxotungstate. Chembiochem. 2015;16(2):233-41. Epub 2014/12/19. doi: 10.1002/cbic.201402597. PubMed PMID: 25521080; PMCID: PMC4498469.

118. Lee CC, Maestre-Reyna M, Hsu KC, Wang HC, Liu CI, Jeng WY, Lin LL, Wood R, Chou CC, Yang JM, Wang AH. Crowning proteins: modulating the protein surface properties using crown ethers. Angew Chem Int Ed Engl. 2014;53(48):13054-8. Epub 2014/10/08. doi: 10.1002/anie.201405664. PubMed PMID: 25287606; PMCID: PMC4288931.

119. Frey L, Lakomek NA, Riek R, Bibow S. Micelles, Bicelles, and Nanodiscs: Comparing the Impact of Membrane Mimetics on Membrane Protein Backbone Dynamics. Angew Chem Int Ed Engl. 2017;56(1):380-3. Epub 2016/11/25. doi: 10.1002/anie.201608246. PubMed PMID: 27882643.

120. Ritchie TK, Grinkova YV, Bayburt TH, Denisov IG, Zolnerciks JK, Atkins WM, Sligar SG. Chapter 11 - Reconstitution of membrane proteins in phospholipid bilayer nanodiscs. Methods Enzymol. 2009;464:211-31. Epub 2009/11/12. doi: 10.1016/S0076-6879(09)64011-8. PubMed PMID: 19903557; PMCID: PMC4196316.

121. Denisov IG, Sligar SG. Nanodiscs for structural and functional studies of membrane proteins. Nat Struct Mol Biol. 2016;23(6):481-6. Epub 2016/06/09. doi: 10.1038/nsmb.3195. PubMed PMID: 27273631.

122. Denisov IG, Grinkova YV, Lazarides AA, Sligar SG. Directed self-assembly of monodisperse phospholipid bilayer Nanodiscs with controlled size. J Am Chem Soc. 2004;126(11):3477-87. Epub 2004/03/18. doi: 10.1021/ja0393574. PubMed PMID: 15025475.

123. Bayburt TH, Grinkova YV, Sligar SG. Self-Assembly of Discoidal Phospholipid Bilayer Nanoparticles with Membrane Scaffold Proteins. Nano Letters. 2002;2(8):853-6. doi: 10.1021/nl025623k.

124. Svergun D, Koch MHJ, Timmins PA, May RP. Small Angle X-ray and Neutron Scattering from Soluitons of Biological Macromolecules2013.

125. Skar-Gislinge N, Simonsen JB, Mortensen K, Feidenhans'l R, Sligar SG, Lindberg Moller B, Bjornholm T, Arleth L. Elliptical structure of phospholipid bilayer nanodiscs encapsulated by scaffold proteins: casting the roles of the lipids and the protein. J Am Chem Soc. 2010;132(39):13713-22. Epub 2010/09/11. doi: 10.1021/ja1030613. PubMed PMID: 20828154; PMCID: PMC4120756.

126. Guinier A. La diffraction de rayons X aux très petits angles: application á l'étude de phénoménes ultramicroscopiques.: Univ. de Paris; 1939.

127. Jacques DA, Trewhella J. Small-angle scattering for structural biology--expanding the frontier while avoiding the pitfalls. Protein Sci. 2010;19(4):642-57. Epub 2010/02/02. doi: 10.1002/pro.351. PubMed PMID: 20120026; PMCID: PMC2867006.

128. Orthaber D, Bergmann A, Glatter O. SAXS experiments on absolute scale with Kratky systems using water as a secondary standard. Journal of Applied Crystallography. 2000;33(2):218-25. doi: doi:10.1107/S0021889899015216.

129. Rambo RP, Tainer JA. Accurate assessment of mass, models and resolution by small-angle scattering. Nature. 2013;496(7446):477-81. Epub 2013/04/27. doi: 10.1038/nature12070. PubMed PMID: 23619693; PMCID: PMC3714217.

130. Mylonas E, Svergun D. Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. Journal of Applied Crystallography. 2007;40:s245-s9.

131. Debnath A, Schafer LV. Structure and Dynamics of Phospholipid Nanodiscs from All-Atom and Coarse-Grained Simulations. J Phys Chem B. 2015;119(23):6991-7002. Epub 2015/05/16. doi: 10.1021/acs.jpcb.5b02101. PubMed PMID: 25978497.

132. Franke D, Petoukhov MV, Konarev PV, Panjkovich A, Tuukkanen A, Mertens HDT, Kikhney AG, Hajizadeh NR, Franklin JM, Jeffries CM, Svergun DI. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. J Appl Crystallogr. 2017;50(Pt 4):1212-25. Epub 2017/08/16. doi: 10.1107/S1600576717007786. PubMed PMID: 28808438; PMCID: PMC5541357.

133. Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. Journal of Applied Crystallography. 2003;36(5):1277-82. doi: doi:10.1107/S0021889803012779.

134. Hayes D, Laue T, Philo J. Sednterp: sedimentation interpretation program. Durham, NH: University of New Hampshire1995.

135. Demeler B. UltraScan: A Comprehensive Data Analysis Software Package for Analytical Ultracentrifugation Experiments. . In: Scott DJ, Harding SE, Rowe AJ, editors. Analytical Ultracentrifugation: Techniques and Methods. UK: Royal Society of Chemistry; 2005. p. 210-29.

136. Demeler B. UltraScan3 Version 4.0 Source Code 2011; Version 3:[Available from: http://www.ultrascan3.uthscsa.edu/doxygen/html/us__math2_8cpp_source.html (lines 473-482).

137. Greenwood AI, Tristram-Nagle S, Nagle JF. Partial molecular volumes of lipids and cholesterol. Chem Phys Lipids. 2006;143(1-2):1-10. Epub 2006/06/02. doi: 10.1016/j.chemphyslip.2006.04.002. PubMed PMID: 16737691; PMCID: PMC2695672.

138. Tristram-Nagle S, Petrache HI, Nagle JF. Structure and interactions of fully hydrated dioleoylphosphatidylcholine bilayers. Biophys J. 1998;75(2):917-25.

Epub 1998/07/24. doi: 10.1016/S0006-3495(98)77580-0. PubMed PMID: 9675192; PMCID: PMC1299765.

139. Schuck P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophys J. 2000;78(3):1606-19. Epub 2000/02/29. doi: 10.1016/S0006-3495(00)76713-0. PubMed PMID: 10692345; PMCID: PMC1300758.

140. Hagn F, Etzkorn M, Raschle T, Wagner G. Optimized phospholipid bilayer nanodiscs facilitate high-resolution structure determination of membrane proteins. J Am Chem Soc. 2013;135(5):1919-25. Epub 2013/01/09. doi: 10.1021/ja310901f. PubMed PMID: 23294159; PMCID: PMC3566289.

141. Kynde SA, Skar-Gislinge N, Pedersen MC, Midtgaard SR, Simonsen JB, Schweins R, Mortensen K, Arleth L. Small-angle scattering gives direct structural information about a membrane protein inside a lipid environment. Acta Crystallogr D Biol Crystallogr. 2014;70(Pt 2):371-83. Epub 2014/02/18. doi: 10.1107/S1399004713028344. PubMed PMID: 24531471.

142. Mazhab-Jafari MT, Marshall CB, Smith MJ, Gasmi-Seabrook GM, Stathopulos PB, Inagaki F, Kay LE, Neel BG, Ikura M. Oncogenic and RASopathy-associated K-RAS mutations relieve membrane-dependent occlusion of the effector-binding site. Proc Natl Acad Sci U S A. 2015;112(21):6625-30. Epub 2015/05/06. doi: 10.1073/pnas.1419895112. PubMed PMID: 25941399; PMCID: PMC4450377.

143. Neuhoff V, Stamm R, Pardowitz I, Arold N, Ehrhardt W, Taube D. Essential problems in quantification of proteins following colloidal staining with coomassie brilliant blue dyes in polyacrylamide gels, and their solution. Electrophoresis. 1990;11(2):101-17. Epub 1990/02/01. doi: 10.1002/elps.1150110202. PubMed PMID: 1692528.

144. Inagaki S, Ghirlando R, Grisshammer R. Biophysical characterization of membrane proteins in nanodiscs. Methods. 2013;59(3):287-300. Epub 2012/12/12. doi: 10.1016/j.ymeth.2012.11.006. PubMed PMID: 23219517; PMCID: PMC3608844.

145. Schuler MA, Denisov IG, Sligar SG. Nanodiscs as a new tool to examine lipid-protein interactions. Methods in molecular biology (Clifton, NJ).

2013;974:415-33. Epub 2013/02/14. doi: 10.1007/978-1-62703-275-9_18. PubMed PMID: 23404286; PMCID: PMC4201044.

146. Kell GS. Isothermal compressibility of liquid water at 1 atm.
. Journal of Chemical and Engineering Data. 1970;15(1):119-22. doi: 10.1021/je60044a003.

147. Wyatt PJ. Measurement of special nanoparticle structures by light scattering. Anal Chem. 2014;86(15):7171-83. Epub 2014/07/23. doi: 10.1021/ac500185w. PubMed PMID: 25047231.

148. Shih AY, Freddolino PL, Sligar SG, Schulten K. Disassembly of nanodiscs with cholate. Nano Lett. 2007;7(6):1692-6. Epub 2007/05/17. doi: 10.1021/nl0706906. PubMed PMID: 17503871.

149. Skar-Gislinge N, Kynde SA, Denisov IG, Ye X, Lenov I, Sligar SG, Arleth L. Small-angle scattering determination of the shape and localization of human cytochrome P450 embedded in a phospholipid nanodisc environment. Acta Crystallogr D Biol Crystallogr. 2015;71(Pt 12):2412-21. Epub 2015/12/03. doi: 10.1107/S1399004715018702. PubMed PMID: 26627649; PMCID: PMC4667284.

150. Franz KJ, Nitz M, Imperiali B. Lanthanide-binding tags as versatile protein coexpression probes. Chembiochem. 2003;4(4):265-71. Epub 2003/04/03. doi: 10.1002/cbic.200390046. PubMed PMID: 12672105.

151. Barthelmes K, Reynolds AM, Peisach E, Jonker HR, DeNunzio NJ, Allen KN, Imperiali B, Schwalbe H. Engineering encodable lanthanide-binding tags into loop regions of proteins. J Am Chem Soc. 2011;133(4):808-19. Epub 2010/12/25. doi: 10.1021/ja104983t. PubMed PMID: 21182275; PMCID: PMC3043167.

152. Martin LJ, Hahnke MJ, Nitz M, Wohnert J, Silvaggi NR, Allen KN, Schwalbe H, Imperiali B. Double-lanthanide-binding tags: design, photophysical properties, and NMR applications. J Am Chem Soc. 2007;129(22):7106-13. Epub 2007/05/15. doi: 10.1021/ja070480v. PubMed PMID: 17497862.

153. Miake-Lye RC, Doniach S, Hodgson KO. Anomalous x-ray scattering from terbium-labeled parvalbumin in solution. Biophys J. 1983;41(3):287-92. Epub 1983/03/01. doi: 10.1016/S0006-3495(83)84440-3. PubMed PMID: 6838970; PMCID: PMC1329182.

154. Stuhrmann HB, Notbohm H. Configuration of the four iron atoms in dissolved human hemoglobin as studied by anomalous dispersion. Proc Natl Acad Sci U S A. 1981;78(10):6216-20. Epub 1981/10/01. PubMed PMID: 6947224; PMCID: PMC349009.

155. Makowski L, Bardhan J, Gore D, Rodi DJ, Fischetti RF. Multi-wavelength anomalous diffraction using medium-angle X-ray solution scattering (MADMAX). Biophys J. 2012;102(4):927-33. Epub 2012/03/06. doi: 10.1016/j.bpj.2012.01.026. PubMed PMID: 22385864; PMCID: PMC3283801.

156. Silvaggi NR, Martin LJ, Schwalbe H, Imperiali B, Allen KN. Double-Lanthanide-Binding Tags for Macromolecular Crystallographic Structure Determination. Journal of the American Chemical Society. 2007;129(22):7114-20. doi: 10.1021/ja070481n.

157. Evans G, Pettifer RF. CHOOCH: a program for deriving anomalous-scattering factors from X-ray fluorescence spectra. Journal of Applied Crystallography. 2001;34(1):82-6. doi: doi:10.1107/S0021889800014655.

158. Svergun D, Barberato C, Koch MHJ. CRYSOL– a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. Journal of Applied Crystallography. 1995;28(6):768-73. doi: 10.1107/s0021889895007047.

159. Cromer DT, Liberman DA. Anomalous dispersion calculations near to and on the long-wavelength side of an absorption edge. Acta Crystallographica Section A. 1981;37(2):267-8. doi: doi:10.1107/S0567739481000600.

160. Peterfy M, Phan J, Xu P, Reue K. Lipodystrophy in the fld mouse results from mutation of a new gene encoding a nuclear protein, lipin. Nat Genet. 2001;27(1):121-4. Epub 2001/01/04. doi: 10.1038/83685. PubMed PMID: 11138012.

161. Csaki LS, Dwyer JR, Fong LG, Tontonoz P, Young SG, Reue K. Lipins, lipinopathies, and the modulation of cellular lipid storage and signaling. Prog Lipid Res. 2013;52(3):305-16. Epub 2013/04/23. doi: 10.1016/j.plipres.2013.04.001. PubMed PMID: 23603613; PMCID: PMC3830937.

162. Reue K, Zhang P. The lipin protein family: dual roles in lipid biosynthesis and gene expression. FEBS Lett. 2008;582(1):90-6. Epub 2007/11/21. doi: 10.1016/j.febslet.2007.11.014. PubMed PMID: 18023282; PMCID: PMC2848953.

163. Bou Khalil M, Blais A, Figeys D, Yao Z. Lipin - The bridge between hepatic glycerolipid biosynthesis and lipoprotein metabolism. Biochim Biophys Acta. 2010;1801(12):1249-59. Epub 2010/08/10. doi: 10.1016/j.bbalip.2010.07.008. PubMed PMID: 20692363.

164. Peterfy M, Phan J, Reue K. Alternatively spliced lipin isoforms exhibit distinct expression pattern, subcellular localization, and role in adipogenesis. J Biol Chem. 2005;280(38):32883-9. Epub 2005/07/29. doi: 10.1074/jbc.M503885200. PubMed PMID: 16049017.

165. Finck BN, Gropler MC, Chen Z, Leone TC, Croce MA, Harris TE, Lawrence JC, Jr., Kelly DP. Lipin 1 is an inducible amplifier of the hepatic PGC-1alpha/PPARalpha regulatory pathway. Cell Metab. 2006;4(3):199-210. Epub 2006/09/05. doi: 10.1016/j.cmet.2006.08.005. PubMed PMID: 16950137.

166. Santos-Rosa H, Leung J, Grimsey N, Peak-Chew S, Siniossoglou S. The yeast lipin Smp2 couples phospholipid biosynthesis to nuclear membrane growth. EMBO J. 2005;24(11):1931-41. Epub 2005/05/13. doi: 10.1038/sj.emboj.7600672. PubMed PMID: 15889145; PMCID: PMC1142606.

167. Kelley LA, Sternberg MJ. Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc. 2009;4(3):363-71. Epub 2009/02/28. doi: 10.1038/nprot.2009.2. PubMed PMID: 19247286.

168. Peterfy M, Harris TE, Fujita N, Reue K. Insulin-stimulated interaction with 14-3-3 promotes cytoplasmic localization of lipin-1 in adipocytes. J Biol Chem. 2010;285(6):3857-64. Epub 2009/12/04. doi: 10.1074/jbc.M109.072488. PubMed PMID: 19955570; PMCID: PMC2823528.

169. Harris TE, Huffman TA, Chi A, Shabanowitz J, Hunt DF, Kumar A, Lawrence JC, Jr. Insulin controls subcellular localization and multisite phosphorylation of the phosphatidic acid phosphatase, lipin 1. J Biol Chem. 2007;282(1):277-86. Epub 2006/11/16. doi: 10.1074/jbc.M609537200. PubMed PMID: 17105729.

170. Kim S, Kedan A, Marom M, Gavert N, Keinan O, Selitrennik M, Laufman O, Lev S. The phosphatidylinositol-transfer protein Nir2 binds phosphatidic acid and positively regulates phosphoinositide signalling. EMBO Rep. 2013;14(10):891-9. Epub 2013/07/31. doi: 10.1038/embor.2013.113. PubMed PMID: 23897088; PMCID: PMC3807235.

171. Carman GM, Han GS. Roles of phosphatidate phosphatase enzymes in lipid metabolism. Trends Biochem Sci. 2006;31(12):694-9. Epub 2006/11/03. doi: 10.1016/j.tibs.2006.10.003. PubMed PMID: 17079146; PMCID: PMC1769311.

172. Park Y, Han GS, Mileykovskaya E, Garrett TA, Carman GM. Altered Lipid Synthesis by Lack of Yeast Pah1 Phosphatidate Phosphatase Reduces Chronological Life Span. J Biol Chem. 2015;290(42):25382-94. Epub 2015/09/05. doi: 10.1074/jbc.M115.680314. PubMed PMID: 26338708; PMCID: PMC4646187.

173. McPherson A, Cudney B. Searching for silver bullets: an alternative strategy for crystallizing macromolecules. J Struct Biol. 2006;156(3):387-406. Epub 2006/11/15. doi: 10.1016/j.jsb.2006.09.006. PubMed PMID: 17101277.

174. Brown PH, Schuck P. Macromolecular size-and-shape distributions by sedimentation velocity analytical ultracentrifugation. Biophys J. 2006;90(12):4651-61. Epub 2006/03/28. doi: 10.1529/biophysj.106.081372. PubMed PMID: 16565040; PMCID: PMC1471869.

175. Weiss JN. The Hill equation revisited: uses and misuses. FASEB J. 1997;11(11):835-41. Epub 1997/09/01. PubMed PMID: 9285481.

176. Ericsson UB, Hallberg BM, Detitta GT, Dekker N, Nordlund P. Thermofluor-based high-throughput stability optimization of proteins for structural studies. Anal Biochem. 2006;357(2):289-98. Epub 2006/09/12. doi: 10.1016/j.ab.2006.07.027. PubMed PMID: 16962548.

177. Hefti MH, Van Vugt-Van der Toorn CJ, Dixon R, Vervoort J. A novel purification method for histidine-tagged proteins containing a thrombin cleavage site. Anal Biochem. 2001;295(2):180-5. Epub 2001/08/08. doi: 10.1006/abio.2001.5214. PubMed PMID: 11488620.

178. Kane JF. Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli. Curr Opin Biotechnol. 1995;6(5):494-500. Epub 1995/10/01. PubMed PMID: 7579660.

179. Joseph RE, Andreotti AH. Bacterial expression and purification of interleukin-2 tyrosine kinase: single step separation of the chaperonin impurity. Protein Expr Purif. 2008;60(2):194-7. Epub 2008/05/23. doi: 10.1016/j.pep.2008.04.001. PubMed PMID: 18495488; PMCID: PMC2581883.

180. Little DJ, Whitney JC, Robinson H, Yip P, Nitz M, Howell PL. Combining in situ proteolysis and mass spectrometry to crystallize Escherichia coli PgaB. Acta Crystallogr Sect F Struct Biol Cryst Commun. 2012;68(Pt 7):842-5. Epub 2012/07/04. doi: 10.1107/S1744309112022075. PubMed PMID: 22750880; PMCID: PMC3388937.

181. Microyltic. Crystal Former User Manual.

182. Luft JR, Wolfley JR, Snell EH. What's in a drop? Correlating observations and outcomes to guide macromolecular crystallization experiments. Cryst Growth Des. 2011;11(3):651-63. Epub 2011/06/07. doi: 10.1021/cg1013945. PubMed PMID: 21643490; PMCID: PMC3106348.

183. Svergun DI. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophys J. 1999;76(6):2879-86. Epub 1999/06/04. doi: 10.1016/S0006-3495(99)77443-6. PubMed PMID: 10354416; PMCID: PMC1300260.

184. Scott DJ, Harding SE, Rowe AJ. Analytical Ultracentrifugation: Techniques and Methods. University of Nottingham, Cambridge: Royal Society of Chemistry; 2005.

185. Medzihradszky KF. Characterization of site-specific N-glycosylation. Methods in molecular biology (Clifton, NJ). 2008;446:293-316. Epub 2008/04/01. doi: 10.1007/978-1-60327-084-7_21. PubMed PMID: 18373266.

186. Chen YZ, Tang YR, Sheng ZY, Zhang Z. Prediction of mucin-type O-glycosylation sites in mammalian proteins using the composition of k-spaced amino acid pairs. BMC Bioinformatics. 2008;9:101. Epub 2008/02/20. doi: 10.1186/1471-2105-9-101. PubMed PMID: 18282281; PMCID: PMC2335299.

187. Zhao H, Brautigam CA, Ghirlando R, Schuck P. Overview of current methods in sedimentation velocity and sedimentation equilibrium analytical ultracentrifugation. Curr Protoc Protein Sci. 2013;Chapter 20:Unit20 12. Epub 2013/02/05. doi: 10.1002/0471140864.ps2012s71. PubMed PMID: 23377850; PMCID: PMC3652391.

188. Seidel SA, Dijkman PM, Lea WA, van den Bogaart G, Jerabek-Willemsen M, Lazic A, Joseph JS, Srinivasan P, Baaske P, Simeonov A, Katritch I, Melo FA, Ladbury JE, Schreiber G, Watts A, Braun D, Duhr S. Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions. Methods. 2013;59(3):301-15. Epub 2012/12/29. doi: 10.1016/j.ymeth.2012.12.005. PubMed PMID: 23270813; PMCID: PMC3644557.

189. Yue L, Ye F, Gui C, Luo H, Cai J, Shen J, Chen K, Shen X, Jiang H. Ligand-binding regulation of LXR/RXR and LXR/PPAR heterodimerizations: SPR technology-based kinetic analysis correlated with molecular dynamics simulation. Protein Sci. 2005;14(3):812-22. Epub 2005/02/22. doi: 10.1110/ps.04951405. PubMed PMID: 15722453; PMCID: PMC2279270.

CURRICULUM VITAE